

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 10, 2004, 10:37:15 ; Search time 763 Seconds  
(without alignments)  
11592.069 Million cell updates/sec

Title: US-10-076-069-1

Perfect score: 2082

Sequence: 1 ccccgcgctgcgctccacc.....atcgaaaaaataaaaaa 2082

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_290a04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002s:\*  
7: geneseqn2003as:\*  
8: geneseqn2003bs:\*  
9: geneseqn2003cs:\*  
10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2082	100.0	2082	6	AA141497
2	428.4	20.6	2038	3	AA1818134
3	428.4	20.6	2038	3	AAH33113
4	426.8	20.5	1314	6	ABL49854
5	426.8	20.5	2082	6	AA141498
6	425.2	20.4	2447	6	AAAD34182
7	425.2	20.4	2629	4	AA160667
8	425.2	20.4	2629	4	AA158881
9	425.2	20.4	2629	8	ADBA48863
10	355	17.1	862	4	AAH06538
11	190	9.1	396	4	AA194976
12	190	9.1	396	6	ABL48926
13	190	9.1	396	6	ABL03243
14	185.4	8.9	522	6	ABK63378
15	98.2	4.7	477	8	ACH24999
16	94.8	4.6	1221	9	ADBD7527
17	62.4	3.0	368	6	ABL80236
18	61.2	2.9	18988	4	AA545342
19	61.2	2.9	18988	6	ABL32701
20	61.2	2.9	18988	6	ABL34509
21	61.2	2.9	18988	6	ABL70204
22	59	2.8	17721	6	ABL33728
23	57.4	2.8	2000	7	ADA71938

24	56.8	2.7	2000	7	ADA71938	ADA71938	Rice gene
25	56.2	2.7	9642	6	ABL32356	ABL32356	Human imm
26	55.8	2.7	883	4	AA15210	AA15210	Human bre
27	55.2	2.7	5689	4	AA545384	AA545384	Chemical
28	55.2	2.7	5689	4	AA546426	AA546426	Tumour su
29	55.2	2.7	5689	6	ABK28226	ABK28226	DNA trans
30	55.2	2.7	83391	6	ABO67093	ABO67093	Human ang
31	54.8	2.6	15548	6	ABL34155	ABL34155	Human imm
32	54.6	2.6	5488	6	ABL33456	ABL33456	Human ang
33	54.4	2.6	33053	6	ABO67006	ABO67006	Human ang
34	54.4	2.6	37973	6	ABL34196	ABL34196	Human imm
35	54	2.6	8136	6	ABK39957	ABK39957	Human che
36	54	2.6	8136	6	ABL32555	ABL32555	Human imm
37	53.8	2.6	7589	6	ABK28390	ABK28390	DNA trans
38	53.6	2.6	113515	6	ABL34174	ABL34174	Human imm
39	53.4	2.6	6128	6	ABL32056	ABL32056	Human imm
40	52.6	2.5	6904	6	ABL32212	ABL32212	Human imm
41	52.2	2.5	5273	6	ABL32875	ABL32875	Human imm
42	52	2.5	15161	6	ABL70457	ABL70457	Chemical
43	52	2.5	15161	6	AA561422	AA561422	Human gen
44	51.8	2.5	6154	6	ABK31201	ABK31201	Signal tr
45	51.8	2.5	6154	6	ABL70168	ABL70168	Chemical

## ALIGNMENTS

RESULT 1	
ID	AA141497
XX	AA141497 standard; DNA; 2082 BP.
XX	
AC	AA141497;
XX	
DT	19-DEC-2002 (first entry)
XX	
DE	Mouse haematopoietic progenitor protein (Hepp) gene.
XX	
KW	Neuroprotective; neurotropic; cytosstatic; neurodegenerative disease; blood;
KW	amyotrophic sclerosis; haematological disorder; neoplasm; leukaemia;
KW	acute myelomonocytic leukaemia; lymphoblastic lymphoma; multiple myeloma;
KW	chronic lymphocytic leukaemia; acute lymphoblastic leukaemia;
KW	B-lymphocytic leukaemia; plasma cell leukaemia; large B-cell lymphoma;
KW	adult T-cell lymphoma; nodal marginal zone B-cell lymphoma; stem cell;
KW	Burkitt's lymphoma; follicular lymphoma; hairy cell leukaemia;
KW	mantle cell lymphoma; splenic marginal zone B-cell lymphoma;
KW	T-lymphocytic leukaemia; haematopoietic cytokine; growth factor;
KW	progenitor cell; gene therapy; Hepp; haematopoietic progenitor protein;
KW	mouse; gene; ds.
XX	
OS	Mus musculus.
XX	
FX	Key
FX	CDS
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	
FT	

PT New hematopoietic progenitor protein (Hep) genes and proteins, useful  
PT for detecting, treating and preventing neurodegenerative diseases, e.g.  
PT amyotrophic sclerosis, and hematological disorders, e.g. neoplasms of the  
PT blood.

PS Claim 5, Fig 1a, 5app; English.

XX  
XX The invention relates to an isolated nucleic acid comprising at least 85%  
CC identity to either of 2 2082 base pair sequences, given in the  
CC specification. The nucleic acids and polypeptides of the invention are  
CC useful for detecting, treating and preventing neurodegenerative diseases  
CC such as amyotrophic sclerosis, and hematological disorders, particularly  
CC neoplasms of the blood such as acute myelomonocytic leukemia,  
CC lymphoblastic lymphoma, chronic lymphocytic leukemia, acute  
CC lymphoblastic leukemia, multiple myeloma, B-prolymphocytic leukemia,  
CC plasma cell leukemia, adult T-cell lymphoma/leukemia, diffuse large B-  
CC cell lymphoma, nodal marginal zone B-cell lymphoma, Burkitt's lymphoma,  
CC follicular lymphoma, hairy cell leukemia, mantle cell lymphoma, splenic  
CC marginal zone B-cell lymphoma, and T-prolymphocytic leukemia. They are  
CC also useful as reagents for differential identification of tissues and  
CC cell types present in the biological sample. The mammal is useful in  
CC screening drugs for treating the disorders cited above, and for testing  
CC of novel hematopoietic cytokines/growth factors for mobilization and  
CC differentiation of stem and progenitor cells. The nucleic acids of the  
CC invention can be used in gene therapy. This polynucleotide sequence  
CC represents the mouse hematopoietic progenitor protein (Hep) gene of the  
CC invention

XX  
XX Sequence 2082 BP; 484 A; 520 C; 522 G; 556 T; 0 U; 0 Other:

Query Match 100.0%; Score 2082; DB 6; Length 2082;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2082; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCCCGCTGGCTTCCACCTCACTTTGAGAGTGGCCGCCCTTCTGTGGCAATTTT 60  
DB 1 CCCCCGCTGGCTTCCACCTCACTTTGAGAGTGGCCGCCCTTCTGTGGCAATTTT 60  
QY 61 GGGGGAGTGGACCTTCCCTGCTTTTAGACGCGCCAGCGCCATGCGACCTTTCTG 120  
DB 61 GGGGGAGTGGACCTTCCCTGCTTTTAGACGCGCCAGCGCCATGCGACCTTTCTG 120  
QY 121 GCGAGTGACCGATTCCGGGTGCCCCGAAGAGCTGGGTGGTCTGCTTTCAGCCGCC 180  
DB 121 GCGAGTGACCGATTCCGGGTGCCCCGAAGAGCTGGGTGGTCTGCTTTCAGCCGCC 180  
QY 181 CCGGACAGAGTGTGCTAGAGGGGCTGAAGAGGAATATGTGACACCAAGAAAGAGT 240  
DB 181 CCGGACAGAGTGTGCTAGAGGGGCTGAAGAGGAATATGTGACACCAAGAAAGAGT 240  
QY 241 AAGAGGTTTGGCACTGCTCTTCTATAGCCTGACGACAGTCACTCTGGAATGTC 300  
DB 241 AAGAGGTTTGGCACTGCTCTTCTATAGCCTGACGACAGTCACTCTGGAATGTC 300  
QY 301 CCTTGTCAAGCTTCAGCTCTGTCAATGCTAGTGAGCCCAATCTTCGCGTCT 360  
DB 301 CCTTGTCAAGCTTCAGCTCTGTCAATGCTAGTGAGCCCAATCTTCGCGTCT 360  
QY 361 CATTGCGCAACAGTCCCGGCAATCCAGAGGAATAGACCAAGATGTGTGTGATG 420  
DB 361 CATTGCGCAACAGTCCCGGCAATCCAGAGGAATAGACCAAGATGTGTGTGATG 420  
QY 421 GATGGACCCCAAGATGTAGATCGGGACACATTTGAACGCTGTGTCCACAGATCT 480  
DB 421 GATGGACCCCAAGATGTAGATCGGGACACATTTGAACGCTGTGTCCACAGATCT 480  
QY 481 GTGTGTACAGTGGAGGAGCTGAGAGAGACACCTGTCTCTGAATCGAAGATGCTCC 540  
DB 481 GTGTGTACAGTGGAGGAGCTGAGAGAGACACCTGTCTCTGAATCGAAGATGCTCC 540  
QY 541 CTGGCAAACTGGTTTCCGAGCTCCCAATCGTTGGCTACGACAGGGGCAAGAAACC 600  
DB 541 CTGGCAAACTGGTTTCCGAGCTCCCAATCGTTGGCTACGACAGGGGCAAGAAACC 600

QY 601 TCAGAGCAGCTCTGGGAGATGACACGCCCAAGAAAACAGGGAGCTTCAGAACTC 660  
DB 601 TCAGAGCAGCTCTGGGAGATGACACGCCCAAGAAAACAGGGAGCTTCAGAACTC 660  
QY 661 ACTGACAGATATTTAGAGACCTTGAGAAACAAAATCTCAGTTGAGGAACTCTT 720  
DB 661 ACTGACAGATATTTAGAGACCTTGAGAAACAAAATCTCAGTTGAGGAACTCTT 720  
QY 721 CTCAGATGTGGACAGTCTCTATATAGCTGTGACACAGTGTCTTAACAGAAATGATG 780  
DB 721 CTCAGATGTGGACAGTCTCTATATAGCTGTGACACAGTGTCTTAACAGAAATGATG 780  
QY 781 GACCAAGTCACTCTCTGAAATGGGCTTGGAGGGCTTGTGAGCCACCCCTCCAG 840  
DB 781 GACCAAGTCACTCTCTGAAATGGGCTTGTGAGGGCTTGTGAGCCACCCCTCCAG 840  
QY 841 TTCCACTTGCAGATGTGACCTGGCTGAGCTGAGACCATGTGTGAGATTTGTGTGAGAC 900  
DB 841 TTCCACTTGCAGATGTGACCTGGCTGAGCTGAGACCATGTGTGAGATTTGTGTGAGAC 900  
QY 901 CTGAGAGGCCACCCAGATGGGCTTAAGGTGAGGCCACAGTCCCATGAGCTCAGTGT 960  
DB 901 CTGAGAGGCCACCCAGATGGGCTTAAGGTGAGGCCACAGTCCCATGAGCTCAGTGT 960  
QY 961 GTTGTGACCCAGAGACAGTAAAGCACTGTCTTAAGAGGGGCTGGCTTTGAGCTCAT 1020  
DB 961 GTTGTGACCCAGAGACAGTAAAGCACTGTCTTAAGAGGGGCTGGCTTTGAGCTCAT 1020  
QY 1021 TATCCTTTTGTGTGACATTTGACATCTGATGTGAGAGATGTGTGACAGTATGTCTAGT 1080  
DB 1021 TATCCTTTTGTGTGACATTTGACATCTGATGTGAGAGATGTGTGACAGTATGTCTAGT 1080  
QY 1081 CTATTTTCAATTAATAGTGAACCTTTTAAATTAATTAATTAATTTTGTGGCAATA 1140  
DB 1081 CTATTTTCAATTAATAGTGAACCTTTTAAATTAATTAATTAATTTTGTGGCAATA 1140  
QY 1141 TTTGTCTTAAGATATATTTTAACTTTTATACCTTTAGATTTTTCAGCTATTTTC 1200  
DB 1141 TTTGTCTTAAGATATATTTTAACTTTTATACCTTTTAACTTTTTCAGCTATTTTC 1200  
QY 1201 TTAATAAGTATATTTTCTACAAACATCTGCTGTCTACATTAAGAAATTTAACT 1260  
DB 1201 TTAATAAGTATATTTTCTACAAACATCTGCTGTCTACATTAAGAAATTTAACT 1260  
QY 1261 AATATCAATTTGTGTGTATTTTAAAGTTTAAATTAAGAAATCTTTTGTACTGATC 1320  
DB 1261 AATATCAATTTGTGTGTATTTTAAAGTTTAAATTAAGAAATCTTTTGTACTGATC 1320  
QY 1321 TCTACACTCCCAAGGCAACTGAATGTGAGCCGGGCTTTTAAATGAGAGGCTCCAG 1380  
DB 1321 TCTACACTCCCAAGGCAACTGAATGTGAGCCGGGCTTTTAAATGAGAGGCTCCAG 1380  
QY 1381 TATGGTCTACATTTAGTAGAGCTTTGAAAAGAACATGACAGCTCACTGCCCTCAC 1440  
DB 1381 TATGGTCTACATTTAGTAGAGCTTTGAAAAGAACATGACAGCTCACTGCCCTCAC 1440  
QY 1441 TGGGTCTGCTGTGGCGGATCGGAGCTCTTCTTCTAGCCCGGTGTGACAGATGCTTTAT 1500  
DB 1441 TGGGTCTGCTGTGGCGGATCGGAGCTCTTCTTCTAGCCCGGTGTGACAGATGCTTTAT 1500  
QY 1501 TATGCTATTTATATATGTAATGACCACTGAAGAGTAACTTATCTCTGGAATCCCAAC 1560  
DB 1501 TATGCTATTTATATATGTAATGACCACTGAAGAGTAACTTATCTCTGGAATCCCAAC 1560  
QY 1561 ACCAGTTCTTCAAGGAGCTGTGTGAGCAAGGCTTATGACAGGCTTGTCTTGGCCATC 1620  
DB 1561 ACCAGTTCTTCAAGGAGCTGTGTGAGCAAGGCTTATGACAGGCTTGTCTTGGCCATC 1620  
QY 1621 ACTGCTGTGTTCCAGGCCAGACATGTGACATGAGAGCATGACATGATCCCGAACACCA 1680  
DB 1621 ACTGCTGTGTTCCAGGCCAGACATGTGACATGAGAGCATGACATGATCCCGAACACCA 1680

Qy	1661	GCACCACTACGCTCCATGTCACGTCAGTGTGTACGTGGAGACCACTGGCTCCAGGACCTGTGCTC	1740
Db	1661	GCACCACTACGCTCCATGTCACGTCAGTGTGTACGTGGAGACCACTGGCTCCAGGACCTGTGCTC	1740
Qy	1741	AGAGAGGGGTGTGCACTGCTCCATCGTGTGCTGGGGGGGACGACGGTGAACCTGTGCTTGCTTGC	1800
Db	1741	AGAGAGGGGTGTGCACTGCTCCATCGTGTGCTGGGGGGGACGACGGTGAACCTGTGCTTGCTTGC	1800
Qy	1801	TTTTAAATGCGTCTTGGACGCTTTTAAAGGTTAAAAACAATCCGACTCCATATGATTTAGG	1860
Db	1801	TTTTAAATGCGTCTTGGACGCTTTTAAAGGTTAAAAACAATCCGACTCCATATGATTTAGG	1860
Qy	1861	GCTCCTCCACCTCTGGGGGTGGGCCCTCATGCTGTCTGCTTGATCTCAAAAGCTTGGTACTC	1920
Db	1861	GCTCCTCCACCTCTGGGGGTGGGCCCTCATGCTGTCTGCTTGATCTCAAAAGCTTGGTACTC	1920
Qy	1921	GGCACTGTCAGACTCCATCCCAATGATCTTTTGTTCCTGTGAGCTTTTGGACTT	1980
Db	1921	GGCACTGTCAGACTCCATCCCAATGATCTTTTGTTCCTGTGAGCTTTTGGACTT	1980
Qy	1981	CCCAACCTGAGCTTAAGGTTTTATTTATATGTGCTTCATATATCAACAATGTAAACCTCA	2040
Db	1981	CCCAACCTGAGCTTAAGGTTTTATTTATATGTGCTTCATATATCAACAATGTAAACCTCA	2040
Qy	2041	CTTTATTTAAAGTATCCAGCAATGGAATGAAAAATGAAAAA	2082
Db	2041	CTTTATTTAAAGTATCCAGCAATGGAATGAAAAATGAAAAA	2082

RESULT 2	
AAFI8134	
ID	AAFI8134 standard; DNA; 2038 BP.

DT 14-MAR-2001 (first entry)

DE Lung cancer associated polynucleotide sequence SEQ ID 153.

KM Human cancer associated protein; neuroprotective; cytosolic;  
KM cardioactive; immunomodulatory; muscular active; vulnary;  
KM gastroenterina; nephrotropic; antiinfective; gynecological;  
KM antibacterial; diagnosis; neural disorder; immune disorder; reproductive;  
KM proliferative disorders; wound healing; infectious diseases; ds.

**Homo sapiens**

PN WO200055180-A2.

PD 21-SEP-2000.

08-MAR-2000; 2000WO-US005918.

PR 12-MAR-1999; 99US-0124270P.

PA (HUMA-) HUMAN GENOME SCI INC.

XX / XXXX

XX 11

DR P-PSDB; AAB58258.

PT Lung cancer

PT such as lung cancer.

PS Claim 1; Page 618; 1425pp; English.

CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer  
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer  
CC associated proteins and polynucleotide sequences, their agonists, and  
CC antagonists may have neuroprotective, cyostatic, cardioactive;

CC immunomodulatory: muscular active general; vulnery; gastrointestinal  
CC general; nephrotoxic; antiinfective; gynecological; or antibacterial  
CC activity. The invention also includes antibodies specific for the protein  
CC or polynucleotide sequences. The lung cancer associated polynucleotide  
CC sequences may be used for detection of lung cancer. chromosome  
CC identification, as chromosome markers, and for numerous other diagnostic  
CC or research purposes. The proteins may be used to treat disorders such as  
CC neural, immune, muscular, reproductive, gastrointestinal, pulmonary,  
CC cardiovascular, renal, and proliferative disorders. The proteins may also  
CC be used in the treatment of wounds and infectious diseases.  
CC Polynucleotide sequences AAF18425 - AAF18433 and peptide AAB58549 are  
CC used in the course of the invention for the identification and  
CC characterisation of the polynucleotide and protein sequences  
XX  
SQ Sequence 2038 BP; 439 A; 529 C; 552 G; 508 T; 0 U; 10 Other;

Query Match	20.6%	Score 428.4	DB 3	Length 2038
Best Local Similarity	74.3%	Pred. No. 9.6e-93		
Matches 559; Conservative	0	Mismatches 181	Indels 12	Gaps 1

Qy	175	GGCCCGCTGGA	CAGATGTTT	GTCTAAGAGGCGCTGAAGAAATATGTGACACGAGAA	234
Db	82	CGCGGGTCA	GACACATATTTTTCACGAGACTGAAGAGAAATGTGTTGGCAGAGAA	141	
Qy	235	AGAGATAGAGGT	-----TTTGGCACTGTCCCTTCTATAGCTTGACGAGACA	282	
Db	142	AGACCTGAGAGGAGCGCTTGGCCGCTTAAGAACAATGTCCTCATACAGCTCAGCGGGGA	201		

283 GTCACCTCCTGGACATGTCCCTTGTCAAGCTCCAGCTCTGTACACATGCTAGTGAGGCCCA 342

Db 202 GTGGCTCTGGACATGTTCTCTGGTGAGTTGCAGCTTGGCCACATGCTTGTGGAGCCCA 261

343 TCTCTGCCGCTCGGTCTCATCGCCAACACACAGTCCGGCAGATCCAGGAGGAATGAGCCA 402

Db 262 TCTGTCCGCTCAGTCTCATGTGCCAACACCGTCCGCGAGATCCAGAAGAGATGACGCA 321

403 GGATGCTGTGGCATGGATGCCACCCAGATGTAGATCGGGCACCACTTGAACGCT 462

Db 322 GGATGGAGCTGGCGCACAGTGGCAACCCAGGCTGCAGAGCGGGCGCCGYTCGACCGCTT 381

463 GGTGTCACAGAGATCCTGTGTCGTACAGTGAGGGGAGCTGAGGAAGAGCACCCCTGCTCC 522

Db 382 GGTCTCCACGAGATCCTGTGCCGTGCAGCGTGGGGCAAGGGGCATCCTGCTCC 441

523 TGAAGTGAAGATGCTCCCTTGCAAACTCGGTTTCCGAGCTCCCATCGTTGGCTCAGC 582

Db 442 TGGCTTGGGGACGGCCACACACAGGTTCCAGTTTCTGACCTTTGCCAGTCACCTCAGC 501

583 ACCAGGCAAGGAACCTCAGAGCAGCCTCTGGAGATGGACAGCCCAAGAAGAAACAG 642

Db 502 ACAGCACCAAGCACCTGCAGAGCAGCGCTGGAGATGATGGCCCTCGAGAAACAG 561

643 GGGAAGCTTTCAGAAGTCACTGGACCAGATATTGAGACCCTGGAGAACAAAACTCCAG 702

Db 562 AGGAGCTTTCACAAGTCACTTGATCAGATATTGAAACGCTGGAGACTAAAAACCCAG 621

QY 703 TTCAGTGAGGAAGTCTTCTCAGATGTGGACAGCTCCTACTATGACCTGGACACAGTGCT 762

Db 622 CTGCATGGAAGAGCTGTTCTCAGACGTGGACAGCCCTACTACGACCTGGACACAGTACT 681

QY 763 AACAGGATGATGAGTGGACCAAGTCCAGTCTCTGCAATGGCCTTGAGGGCTTGTCTGC 822

Db 682 GACAGGCATGATGGGGGTGCCAGGCCGGGCCCTGCGAAGGCTCGAGGGCTTGGCTCC 741

823 AGCCACCCCTCTCCCAAGTTCACCTGCAAGTCTGACCTGGCTGAGCTGACCATGTGT 882

Db 742 GGCCACCCCGGGCCTAGCTCCAGCTGCAAGTCCGACCTGGCGGAGCTGGAACCAAGTGT 801

QY 883 AGAGATTCTGGTGAGACCTGAGAGGCCACCC 914

Db 802 GGAGATCCTGGTGGAGACCTGAGCAGGAGCCC 833

```
RESULT 3
AAH33113
ID AAH33113 standard; cDNA; 2038 BP.
AC AAH33113;
XX
XX
XX 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen encoding cDNA SEQ ID NO:169.
XX
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma; ss.
XX
XX Homo sapiens.
XX
XX W0200122920-A2.
XX
XX 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US026524.
XX
XX 29-SEP-1999; 99US-0157137P.
XX 03-NOV-1999; 99US-0163280P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
XX WPI: 2001-235357/24.
XX P-PSDB: AAG73682.
XX
XX Nucleic acid encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers.
XX
XX Claim 1; Page 2326; 9803p; English.
XX
XX AAH32943 to AAH7195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where the
XX proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene therapy
XX and vaccine production. N and P may be used in the prevention, diagnosis
XX and treatment of diseases associated with inappropriate P expression. For
XX example, N and P may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of P by expressing inactive proteins or to
XX supplement the patients own production of P. Additionally, N may be used
XX to produce the colon cancer-associated Ps, by inserting the nucleic acids
XX into a host cell and culturing the cell to express the proteins. N and P
XX can be used in the prevention, diagnosis and treatment of colorectal
XX carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent
XX sequences used in the exemplification of the present invention. N.B.
XX Pages 666 to 682 and page 7053 of the sequence listing were missing at
XX time of publication, meaning no sequences are present for SEQ ID NO:1027
XX to 1052, 7921 and 7922
XX
XX Sequence 2038 BP; 439 A; 529 C; 552 G; 508 T; 0 U; 10 Other;
XX
XX Query Match 20.6%; Score 428.4; DB 4; Length 2038;
XX Best Local Similarity 74.3%; Pred. No. 9.6e-93;
XX Matches 559; Conservative 0; Mismatches 181; Indels 12; Gaps 1;
XX
XX
XX 175 CGCGCCCTGGAAGAGATGTTTCTAGAGGCTGAAGAGAAATATGTGACCAAGAGA 234
XX DB 82 CGGGGTGACAGACAAATGTTTGACAGAGACTGAAGAGAAATGTGTGGCCACAGAGA 141
XX
XX 235 AGGAGTAGAGGCT-----TTTGGCACTGACCTTCTCTAAGCTCAGAGAGAA 282
XX DB 142 AAGACGTGAGAGGAGCCCTGCGCGCTTGAAGACAGTGTCTCAATACGCTCAGAGGCA 201
XX
XX 283 GTCACTCTGAGACATGCTCTTGTCAAGCTCAGCTCTGTCAATGCTAGTGAAGCCCA 342
XX DB 202 GTGCTCTCTGACATGTCTGTGTGAAGTTGACAGCTTTGCCACATGCTGTGAGGCCCA 261
XX
```

```
QY 343 TCTCTGCCGCTCGGTCTCTCATGCGCAACAGTCCCGCATGTCAGAGAAATGACCA 402
DB 262 TCTGTCCGCTCATGCTCTCATGCGCAACAGTCCCGCATGTCAGAGAAATGACCA 321
QY 403 GGATGCTGTGTGACATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 462
DB 322 GGATGAGAGTGTGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 381
QY 463 GGTGTCCACAGAGATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 522
DB 382 GGTCTCCAGGAGATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 441
QY 523 TGAACCTGGAAGATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 582
DB 442 TGGCTTGGGGAGAGGCGCACACACAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 501
QY 583 ACCAGGGCAAGAAAGAAACCTCTCAAGAGACCTCTGAGAGATGACACCCACAGAAACAG 642
DB 502 ACAGGACCAAGGACACTGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 561
QY 643 GGGAGCTTTTCAAGAGTCACTGAGACCAAGATATTGAGACCTGAGAAACAAAACTCGAG 702
DB 562 AGGAGCTTTTCAAGAGTCACTGAGACCAAGATATTGAGACCTGAGAGACTAAACCCAG 621
QY 703 TTCAAGTGAAGAACTCTTCTGAGATGTGAGACAGCTCTCAATATGACCTGAGACAGTCT 762
DB 622 CTGATGAGAAAGCTGTTTCTGAGACAGTGTGAGACAGCTCTCAATATGACCTGAGACAGT 681
QY 763 AACAGAAATGATGAGTGGAGACCAAGTCTGCAATGAGCTTGTGAGGCTTGTGCTGC 822
DB 682 GACAGGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 741
QY 823 AGCCACCCCTCTCTCCAGTCTGCAATGCTGCAATGCTGCAATGCTGCAATGCTGCAAT 882
DB 742 GGCACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 801
QY 883 AGAGATCTGTGTGAGACCTGAGAGGCCACCC 914
DB 802 GGAGATCTGTGTGAGACCTGAGAGGCCACCC 833
XX
XX RESULT 4
XX ABL49854
XX ID ABL49854 standard; cDNA; 1314 BP.
XX
XX ABL49854;
XX
XX 05-JUN-2002 (first entry)
XX
XX Human CHD protein 26.51 encoding cDNA SEQ ID NO:1.
XX
XX Human; CHD protein 26.51; malignant tumour; haemopathy; HIV infection;
XX development disturbance; immunological disease; inflammation; gene; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 98..823
XX FT /*tag= a
XX FT /product= "Human CHD protein 26.51"
XX
XX CNI324856-A.
XX
XX 05-DEC-2001.
XX
XX 24-MAY-2000; 2000CN-00115839.
XX
XX 24-MAY-2000; 2000CN-00115839.
XX
XX (BODE-) BODE GENB DEV CO LTD SHANGHAI.
XX
XX Mao Y, Xie Y;
XX
```

DR WPI; 2002-217513/28.  
 DR P-PSDB; ABB06375.  
 XX  
 PT New polypeptide human CHD protein 26.51 and polynucleotides for encoding same.  
 PS  
 XX Claim 6; Page 25-26 (Disclosure); 33pp; Chinese.  
 CC The present sequence encodes human CHD protein 26.51 (I). The present invention also describes a method for producing (I) using DNA recombination techniques. (I) and the polynucleotide encoding it can be used in the treatment of several diseases, such as malignant tumour, CC haemopathy, development disturbances, HIV infection, immunological CC diseases and various inflammations  
 XX  
 SQ Sequence 1314 BP; 264 A; 376 C; 386 G; 288 T; 0 U; 0 Other;  
 Query Match 20.5%; Score 426.8; DB 6; Length 1314;  
 Best Local Similarity 74.2%; Pred. No. 1.9e-92;  
 Matches 558; Conservative 0; Mismatches 182; Indels 12; Gaps 1;  
 QY 175 CGCCCGCTGACAGATGTTGCTAGAGGCTGAGAGAAATATGTCACCGAGAGA 234  
 DB CGCGGGTCAGAGACACATGTTGACAGAGACTGAAAGAAATGTTGGCCACGAGGA 141  
 QY 235 AGAGATGAGAGGT-----TTTGACATGTCCTTCCATATGCTGACGAGCA 282  
 DB AGACGTGAGAGAGAGCCCTGCGCGGCTTGAAGACATGTTCTCTATACAGCCCTGACGCGCA 201  
 QY 283 GTCACTCTGAGACATGTCCTTGTCAAGCTTCAGACTGTACATGCTAGTGAAGCCAA 342  
 DB 202 GTGCTCTCTGACATGTCCTTGTGAGAGTTCGACCTTTCACATGCTTGTGAGAGCCAA 261  
 QY 343 TCTCTGCGCTCGGTCTCTATCGCAACACAGTCCGCAAGATCCAGAGAAATGAGCA 402  
 DB 262 CTTGTGCGCTCAAGCTCTCAATGCGCAACAGGTCGCAAGAGAGGAGATGAGAGCA 321  
 QY 403 GGATGATGATGGGATGGAGATGGACCCGAAATGATGATCCGGGACACAGTGAAGCCCT 462  
 DB 322 GGATGAGAGTGGGACAGATGGACCCGAGCTGACAGAGGCGCGCTCGACCCCTT 381  
 QY 463 GGTGTCCACAGATCTCTGTGTCTATACATGAGGGAGCTGAGAGAGAGCACTCTGCTCC 522  
 DB 382 GGTCTCCAGGAGATCTGTGTGCGGTGACAGGTGGGGGCAAGAGGGGCACTCTGCTCC 441  
 QY 523 TGAATGGAAGATGCTCTTGCAGAACTCGGTTTCCGAGCTCCCATCTGTGCTCAGC 582  
 DB 442 TGGCTTGGGGAGCGGCGACACACAGGGGTCCAGTTTCTGACCTTGGCCAGTCACTCAGC 501  
 QY 583 ACCAGGCAAGAGACCTTCAGAGCAAGCTCTGGGAGATGAGACGCCCAAGAAAAAG 642  
 DB 502 ACAGGCAACAGGACCTGACAGAGCAGCGCTGGGAGATGATGCTTCCGAGAAAAAG 561  
 QY 643 GGGAGGCTTTCAGAAATGCTGACAGTATTTGAGACCTTGAGAAACAAACTCCAG 702  
 DB 562 AGGAAGTTTCAAGTCACTTGAATGATTTGAAACCTGAGAGCTTAAAAACCCAG 621  
 QY 703 TTCAGTGAAGAACTTCTCAATGATGAGACAGCTCTCTACTATGACCTGAGACAGTGT 762  
 DB 622 CTGATGAGAGAGCTGTTCTCAAGCGTGAAGAGCCCTTACTGACCTGAGACAGTACT 681  
 QY 763 AACAGGAATATGATGAGGAGCAAGTCACTTTCAGAAATGAGCTTGAAGGCTTGTGTC 822  
 DB 682 GACAGGATATAGGGGGGTGCGAGGCGCGGCGCTTGGAAAGGCTTCAGAGGCTTGGCTCC 741  
 QY 823 AGCCACCCCTCTCCAGTTCACACTTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 882  
 DB 742 GGCACCCCAAGGCTTACCTGCTCAGCTCAAGTCCGACCTGGGCAAGCTTGAACACGCTGT 801  
 QY 883 AGAGATTTGATGAGACTGAGAGGCAAGCC 914  
 DB 802 GGAGATCTGTGAGAGCTGAGAGGAGGCC 833

RESULT 5  
 AAL41498  
 ID AAL41498 standard; DNA; 2082 BP.  
 AC  
 XX  
 AC AAL41498;  
 XX  
 DT 19-DEC-2002 (first entry)  
 XX  
 DE Human haematopoietic progenitor protein (Hepp) gene.  
 XX  
 KW Neuroprotective; nootropic; cytoskeletal; neurodegenerative disease; blood;  
 KW amyotrophic sclerosis; haematological disorder; neoplasm; leukaemia;  
 KW acute myelomonocytic leukaemia; lymphoblastic lymphoma; multiple myeloma;  
 KW chronic lymphocytic leukaemia; acute lymphoblastic leukaemia;  
 KW B-prolymphocytic leukaemia; plasma cell leukaemia; large B-cell lymphoma;  
 KW adult T-cell lymphoma; nodal marginal zone B-cell lymphoma; stem cell;  
 KW Burkitt's lymphoma; follicular lymphoma; hairy cell leukaemia;  
 KW mantle cell lymphoma; splenic marginal zone B-cell lymphoma;  
 KW T-prolymphocytic leukaemia; haematopoietic cytokine; growth factor;  
 KW progenitor cell; gene therapy; Hepp; haematopoietic progenitor protein;  
 KW human; gene; ds.  
 XX  
 XX Homo sapiens.  
 XX  
 FT Key Location/Qualifiers  
 FT CDS 75..800  
 FT /\*tag= a  
 FT /product= "Human Hepp protein"  
 FT polyA\_signal 2043..2047  
 FT /\*tag= b  
 XX  
 PN NO20026610-A2.  
 XX  
 XX 29-AUG-2002.  
 PD  
 XX  
 XX 15-FEB-2002; 2002MO-US004503.  
 PF  
 XX  
 XX 16-FEB-2001; 2001US-0268923P.  
 PR  
 XX  
 XX (UWI-) UNIV MIAMI.  
 PA  
 XX  
 XX Juretic R, Nachtmann RG;  
 PI  
 XX  
 XX WPI; 2002-674928/72.  
 DR  
 DR P-PSDB; AAO22898.  
 XX  
 XX  
 PT New haematopoietic progenitor protein (Hepp) genes and proteins, useful  
 PT for detecting, treating and preventing neurodegenerative diseases, e.g.  
 PT amyotrophic sclerosis, and hematological disorders, e.g. neoplasms of the  
 PT blood.  
 PS  
 XX Claim 5; Fig 1B; 54pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising at least 85%  
 CC identity to either of 2 2082 base pair sequences, given in the  
 CC specification. The nucleic acids and polypeptides of the invention are  
 CC useful for detecting, treating and preventing neurodegenerative diseases  
 CC such as amyotrophic sclerosis, and haematological disorders, particularly  
 CC neoplasms of the blood such as acute myelomonocytic leukaemia,  
 CC lymphoblastic lymphoma, chronic lymphocytic leukaemia, acute  
 CC lymphoblastic leukaemia, multiple myeloma, B-prolymphocytic leukaemia,  
 CC plasma cell leukaemia, adult T-cell lymphoma/leukaemia, diffuse large B-  
 CC cell lymphoma, nodal marginal zone B-cell lymphoma, Burkitt's lymphoma,  
 CC follicular lymphoma, hairy cell leukaemia, mantle cell lymphoma, splenic  
 CC marginal zone B-cell lymphoma, and T-prolymphocytic leukaemia. They are  
 CC also useful as reagents for differential identification of tissues and  
 CC cell types present in the biological sample. The mammal is useful in  
 CC screening drugs for treating the disorders cited above, and for testing  
 CC of novel haematopoietic cytokines/growth factors for mobilisation and  
 CC differentiation of stem and progenitor cells. The nucleic acids of the  
 CC invention can be used in gene therapy. This polynucleotide sequence  
 CC represents the human haematopoietic progenitor protein (Hepp) gene of the

CC invention  
XX Sequence 2082 BP; 471 A; 539 C; 545 G; 527 T; 0 U; 0 Other;  
SQ

Query Match 20.5%; Score 426.8; DB 6; Length 2082;  
Best Local Similarity 74.2%; Pred. No. 2.4e-92;  
Matches 558; Conservative 0; Mismatches 182; Indels 12; Gaps 1;

QY 175 CGCCGCTGGAGAGATGTTGTAGAGGGCTGAAGGAAATATGTGTACCGAGAGA 234  
DB 59 CGCGGGTAGAGACACATGTTTGCACGAGACTGAAGGAAATGTGTGGCCACGAGGA 118  
QY 235 AGAGATAGAGGCT-----TTTGGCACTGTCCCTCTCTAATAGCCTGACGACA 282  
DB 119 AGAGGTGAGGAGAGCCTTGCCGGCTTGAAGACAGTGTCTCTAATAGCCTGACGAGCA 178  
QY 283 GTCACTCTGAGACATGTCTCTTGTCAAGCTCCAGCTGTGTCAATGCTAGTGAGCCAA 342  
DB 179 GTCCCTCTGAGACATGTCTCTGTGAAGTTGACGCTTTGACACATGCTGTGTGAGCCAA 238  
QY 343 TCTGCGCGCTCGGCTCTCATGCGCAACAGCTCGGAGATCCAGAGAGAAATGAGCA 402  
DB 239 CTTGTGCGCTCATGCTCTCTCATTTGCCACACGCTCGGAGATCCAGAGAGATGACGA 238  
QY 403 GGATGTGTGTGGCATGGGATGGACCCGCAATGTAGATGGGACACGATGAAAGCCT 462  
DB 299 GGATGGGACGTGGCGCACAGTGCGCACCCGAGGCTGCAAGAGCGGCGCGCTCGACGCTT 358  
QY 463 GGCTGCAACAGATCTCTGTCTCTGTACATGAGGGGAGCTGAGGAAAGCACTCTGCTCC 522  
DB 359 GGTCTTCCAGGAGATCTCTGTGCGCTGTGACGCTGGGGCAAGAGGGGCAACATCTGCTCC 418  
QY 523 TGAATCTGAAGATCTCTCTTGCANAACTCGCTTTCGAGCTCCCATCTGTGGCTGAGC 582  
DB 419 TGGCTTGGGGGACGCGCACACAGGGTCCAGTTTTCGACTTTGGCCAGTCACTTACG 478  
QY 583 ACCAGGGCAAGAGACCTTCAAGACAGCTCTTGGAGATGACAGCCACAGAGAAACAG 642  
DB 479 ACAGGCAACAGGACCTTGACAGACAGCGCTGGGAGATGGAGCTCGAGAGAAACAG 538  
QY 643 GGGAGCTTTCAGAACTCAGTGGACCGAGATATTGAAACCTTGAGAAACAAATCTCCAG 702  
DB 539 AGGAAGTTTTCAGAGCTTGTATTCAGATATTGAAACCTTGAGAACTTAAACCCAG 598  
QY 703 TTCACTGAGAGAACTTCTCAGATGTGACAGCTCTTCACTATGACCTGAGACAGTGTCT 762  
DB 599 CTGATGAGAGAGCTGTCTTCAAGCGTGAACAGCCCTTAATAGACTGAGACAGTACT 658  
QY 763 AACAGGAATATGATGAGGAGCAAGTCAAGTCTCTGCAATGAGCTTGAAGGCTTGTCTGC 822  
DB 659 GACAGGCAATGATGGGGGCTGCCAGGCGGCGCCCTTGGAAGGAGCTGAGAGGCTTGGCTCC 718  
QY 823 AGCCACCCCTCTCTCCAGTTCACCTTGAAGTCTGACCTGGAGCTGAGACCATGTGT 882  
DB 719 GGGCACTCCAGGCTTACCTTCAAGCTGCAAGTCCGACCTGGGAGAGCTGAGACCATGTGT 778  
QY 883 AGAGATTTGTGTGAGAGACCTGAGAGGCAACCC 914  
DB 779 GGAGATCTGTGTGAGAGCTGAGAGAGAGCC 810

RESULT 6  
AAD34182  
ID AAD34182 standard; cDNA; 2447 BP.  
XX  
XX  
AC AAD34182;  
XX  
DT 16-JUL-2002 (first entry)  
XX  
DE Human cDNA coexpressed along with cyclin B and UbcH10 cell cycle gene.  
XX  
XX Human; cell cycle gene; topo II; PRC1; CDC23; cyclin B; UbcH10; hp55cd;  
KM cancer; immune disorder; adenocarcinoma; leukaemia; lymphoma; melanoma;

KM myeloma; sarcoma; atherosclerosis; Crohn's disease; glomerulonephritis;  
KM multiple sclerosis; asthma; myasthenia gravis; rheumatoid arthritis;  
KM scleroderma; osteoporosis; systemic lupus erythematosus; cytostatic;  
KM nephrotropic; neuroprotective; dermatological; immunosuppressive;  
KM osteopathic; ss.  
OS Homo sapiens.  
XX  
XX WO200218575-A2.  
XX  
XX 07-MAR-2002.  
XX  
XX 27-AUG-2001; 2001WO-US026682.  
XX  
XX 30-AUG-2000; 2000US-0229253P.  
XX  
XX (INCY-) INCYTE GENOMICS INC.  
XX Walker MG, Jung K;  
XX  
XX WPI; 2002-329773/36.  
XX  
XX  
XX Composition comprising cDNA molecules coexpressed with one or more known  
PT cell cycle genes, useful for diagnosis and treatment of cell cycle  
PT disorders e.g. glomerulonephritis, multiple sclerosis, rheumatoid  
PT arthritis.  
XX  
XX Claim 6; Page 38-39; 43pp; English.  
XX  
XX The present invention relates to compositions comprising cDNA molecules  
CC coexpressed with one or more known cell cycle genes such as topo II,  
CC PRC1, CDC23, cyclin B, UbcH10 and hp55cd. The composition is useful for  
CC screening several molecules or compounds such as DNA or RNA molecules,  
CC peptide nucleic acids, enhancers, transcription factors, repressors,  
CC mimetics or proteins, identify an purify a ligand. Sequences of the  
CC invention are useful for treating cell cycle disorders such as cancer or  
CC immune disorder such as adenocarcinoma, leukaemia, lymphoma, melanoma,  
CC myeloma, sarcoma or cancers of the blood or bone, asthma,  
CC atherosclerosis, Crohn's disease, glomerulonephritis, multiple sclerosis,  
CC myasthenia gravis, osteoporosis, rheumatoid arthritis, scleroderma or  
CC systemic lupus erythematosus. They are useful as surrogate markers in  
CC diagnosis, prognosis and evaluation of therapies for cell cycle disorders  
CC and potentially serve as the therapeutic for the elimination or control  
CC of unregulated cell cycling. The proteins or peptides expressed from the  
CC cDNAs of the invention are either potential therapeutic or targets for  
CC identification or development of therapeutics. The present sequence is a  
CC cDNA coexpressed along with cyclin B and UbcH10 cell cycle gene. This  
CC cDNA is used as a specific diagnostic marker for breast cancer  
XX  
SQ Sequence 2447 BP; 538 A; 623 C; 638 G; 648 T; 0 U; 0 Other;

Query Match 20.4%; Score 425.2; DB 6; Length 2447;  
Best Local Similarity 74.1%; Pred. No. 6.1e-92;  
Matches 557; Conservative 0; Mismatches 183; Indels 12; Gaps 1;

QY 175 CGCCGCTGGAGAGATGTTGTAGAGGGCTGAAGGAAATATGTGTACCGAGAGA 234  
DB 75 CGCGGGTAGAGACACATGTTTGCACGAGACTGAAGGAAATGTGTGGCCACGAGGA 134  
QY 235 AGAGATAGAGGCT-----TTTGGCACTGTCCCTCTCTAATAGCCTGACGAGCA 282  
DB 135 AGAGGTGAGGAGAGCCTTGCCGGCTTGAAGACAGTGTCTCTAATAGCCTGACGAGCA 194  
QY 283 GTCACTCTGAGACATGTCTCTTGTCAAGCTCCAGCTGTGTCAATGCTAGTGAGCCAA 342  
DB 195 GTCCCTCTGAGACATGTCTCTGTGAAGTTGACAGCTTTGACACATGCTTGTGAGCCAA 254  
QY 343 TCTGCGCGCTCGGCTCTCATGCGCAACAGTCCGAGATCCAGAGAGAAATGAGCCA 402  
DB 255 CTTGTGCGCTCATGCTCTCTATGCGCAACAGGCTCCGACATCCAGAGAGATGAGGCA 314  
QY 403 GGATGTGTGTGGCATGGGATGGACCCGCAAAATGTAGATCGGAGCAACGATGAAAGCCTT 462

Db 315 GGAATGGACGTGGCGCAGATGGCAACCCAGGCTGACAGAGGGCGCCGCTGCACCCCTT 374  
Qy 463 GGTGTCCACAGAGATCTCTGTGTCTGTACAGTAGAGGGGAGCTGAGGAAGAAGCACTCTGTCTCC 522  
Db 375 GGTCTCCACGAGATCTGTGTCTGTGTGACGCTGGGGGCAAGAGGGGGCAACATCTGTCTCC 434  
Qy 523 TGAACGTGAAGATCTCTCTGCAAAAATCGGTTTCGAGCTCCCAATCGTTGCTCAGC 582  
Db 435 TGGCTTGGGGGACGGCCACACACAGGGGTCCAGTTTCTGACCTTTGGCCAGTCACTCAGC 494  
Qy 583 ACCAGGGCAAAAGAACCTCTGAGACAGCTCTGGAGATGAGACGCCCAAGAAAAAG 642  
Db 495 ACAGGGCAACCAAGGACCTGAGACAGACGCTGGAGATGGATGGCCCTCGAGAAAAAG 554  
Qy 643 GGAAGCTTTCAGAAAGTCACTGAGCCAGATTTTGAACCTTGAGAAACAAAACTCAG 702  
Db 555 AGAAGCTTTCACAAAGTCACTGATGATTTGAAACGCTGAGACTAAAAACCCAG 614  
Qy 703 TTCAGTGGAGAACTCTTCTCAGATGTGACAGCTCTACTATGACCTGAGACAGTGT 762  
Db 615 CTGATGGAAGAGCTGTCTTCTCAGACGTGAGACGCCCTTACTACACCTGAGACAGTACT 674  
Qy 763 AACAGATATATGATGTGGAGCAAAATGTCAGTCTCTGCAATGGCCTTGAAGGCTTTGCTGC 822  
Db 675 GACAGGATATGAGGGGGGTGCGACGGCGCCCTCGAAAGGGCTCGAGGGCTTGGCTCC 734  
Qy 823 AGCCACCCCTCTCTCCAGTTTCCACTTGAAGTCTGACCTGGCTGAGCTGAGCCATGTGT 882  
Db 735 GGGCACCACCAAGGCTTGTGCTTCAGCTGCAAGTCCGACCTGGCGAGCTGAGCAAGTAT 794  
Qy 883 AGAGATCTGTGAGAGACTGAGAGGCCACCC 914  
Db 795 GGAGATCTGTGTGAGACTGAGAGAGAGCC 826

## RESULT 7

AAI60667  
ID AAI60667 standard; cDNA; 2629 BP.

AC AAI60667;

DT 22-OCT-2001 (first entry)

DE Human polynucleotide SEQ ID NO 4656.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KM peripheral nervous system; neuropathy; central nervous system; CNS;  
KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KM chemokine; thrombolytic; drug screening; arthritis; inflammation;  
XX leukemia; ss.

XX Homo sapiens.

XX MO20015312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000MO-US034263.

XX 23-DEC-1999; 99US-00471275.

XX 21-JAN-2000; 2000US-00488725.

XX 25-APR-2000; 2000US-00552317.

XX 20-JUN-2000; 2000US-00598042.

XX 19-JUL-2000; 2000US-00620312.

XX 03-AUG-2000; 2000US-00653450.

XX 14-SEP-2000; 2000US-00662191.

XX 19-OCT-2000; 2000US-00693036.

XX 29-NOV-2000; 2000US-00727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
PI Zhou P, Goodrich R, Drmanac RT;  
XX WPI; 2001-442253/47.  
XX DR P-PSDB; AAM41511.  
XX PT Novel nucleic acids and polypeptides, useful for treating disorders such  
XX as central nervous system injuries.  
XX PS Claim 1; SEQ ID NO 4656; 10078pp; English.  
XX The invention relates to human nucleic acids (AA157798-AA161369) and the  
XX encoded polypeptides (AAM38642-AA42213) with nootropic,  
XX immunosuppressant and cytostatic activity. The polynucleotides are useful  
XX in gene therapy. A composition containing a polypeptide or polynucleotide  
XX of the invention may be used to treat diseases of the peripheral nervous  
XX system, such as peripheral nervous injuries, peripheral neuropathy and  
XX localised neuropathies and central nervous system diseases, such as  
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
XX utilisation of the activities such as: Immune system suppression,  
XX Activin/inhibin activity, chemocactive/chemokinetic activity, haemostatic  
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
XX assays for receptor activity, arthritis and inflammation, leukaemia and  
XX C.N.S disorders. Note: The sequence data for this patent did not form  
XX part of the printed specification

XX Sequence 2629 BP; 581 A; 652 C; 691 G; 705 T; 0 U; 0 Other;

XX Query Match 20.4%; Score 425.2; DB 4; Length 2629;

XX Best Local Similarity 74.1%; Pred. No. 6.3e-92;

XX Matches 557; Conservative 0; Mismatches 183; Indels 12; Gaps 1;

Qy 175 CGCCCGCTGACAGATGTTTCTAGAGGGCTGAGAGGAAATATGTGACCAAGAGA 234  
Db 57 CGCGGTGACAGACAAATGTTTGACAGAGACTGAGAGGAAATGTTGGCCACGAGA 116  
Qy 235 AGAGATGAGAGGT-----TTGGCACTGTCCCTCTATGACCTGAGCGACA 282  
Db 117 AGAGGTGAGAGAGCCCTGCGCGCTTGAAGACAGTCTCTATACGCTGACGCGCA 176  
Qy 283 GTCACTCTGAGATGTCCTTGTCAAGCTCCAGCTCTGACATGCTAGTGAAGCCCA 342  
Db 177 GTGCTCTGAGATGTCCTTGTGAAGTGTGACGCTTGGCAATGCTTGTGAAGCCCA 236  
Qy 343 TCTTGTGCGCTGCTCTCTATGCGCAACACAGTCCGCGAGATCCAGAGAAATGAGCA 402  
Db 237 CTGTGTGCGCTCAATCTCTATGCGCAACAGTCCGCGAGATCCAGAGAGATGAGCA 296  
Qy 403 GATGTGTGTGCAATGAGATGAGCAACCCCAAGATATGATGCGGACCAATGAAGCCT 462  
Db 297 GATGTGATGCTGCGCAAGATGAGCAACCCCAAGATATGATGCGGACCAATGAAGCCT 356  
Qy 463 GGTGTCCACAGATCTGTGTGTGTACAGTGAAGGAGAGTGAAGAAAGACCTGTCTCC 522  
Db 357 GGTCTCCACGAGATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 416  
Qy 523 TGAATGGAAGATCTCTTTCGAAAACTGCGTTTCGAGCTCCCAATGCTGTGCTCAGC 582  
Db 417 TGGCTTGGGGGACGGCCACACAGAGGTTCAGATTTTCGACTTTGGCCAGTCACTCAGC 476  
Qy 583 ACCAGGGCAAAAGAACCTCTGAGAGCGCTCTGGAGATGAGACAGCCCAAGAAAAAG 642  
Db 477 ACAGGGCAACCAAGGACCTGAGAGCGCTCTGGAGATGAGAGCCCTGAGAAAAAG 536  
Qy 643 GGAAGCTTTCAGAAAGTCACTGAGACAGATATTTGAACCTGAGAAACAAACTCAG 702  
Db 537 AGAAGCTTTCAGAAAGTCACTGATGATTTTGAAGCTGAGAACTAAAAACCCAG 596  
Qy 703 TTCAGTGAAGAACTCTTCTCAGATGTGACAGCTCTACTATGACCTGAGACAGTGT 762  
Db 597 CTGATGGAAGAGCTGTCTCAGAGTGAAGACCCCTACTAGACTGAGACAGTACT 656





XX	ss; cancer; neurodegenerative disease; human.
KW	
XX	
OS	Homo sapiens.
XX	
FN	US2003104529-A1.
XX	
PD	05-JUN-2003.
XX	
PF	04-JAN-2002; 2002US-00037270.
XX	
PR	21-JAN-2000; 2000US-00488725.
PR	25-APR-2000; 2000US-00552317.
PR	19-JUL-2000; 2000US-00620312.
XX	
PA	(ZHOU/) ZHOU P.
PA	(TANG/) TANG Y T.
PA	(LIUC/) LIU C.
PA	(ASUN/) ASUNDI V.
PA	(DRMA/) DRMANAC R T.
XX	
PI	Zhou P, Tang YT, Liu C, Asundi V, Drmanac RT;
XX	
XX	WPI; 2003-678194/64.
XX	

**PT** New polynucleotide, useful for treating diseases e.g., cancer or neurodegenerative diseases.

PS Claim 1; SEQ ID NO 773; 99pp; English.

The invention relates to a polynucleotide comprising a sequence given in the sequence specification, or its mature protein-coding portion, or its complement. The polynucleotide is useful for treating diseases e.g., cancer or neurodegenerative diseases and many others listed in the sequence specification. The present sequence represents a novel human cDNA. Note: The sequence data for this patent did not form part of the printed publication but was obtained in electronic format directly from USPTO at [seqdata.uspto.gov/sequence.html?DocID=20030104523](http://seqdata.uspto.gov/sequence.html?DocID=20030104523).

**SQ** Sequence 2629 BP; 581 A; 652 C; 691 G; 705 T; 0 U; 0 Other;

Query Match	20.4%	Score	425.2;	DB	8;	Length	2629;
Best Local Similarity	74.1%;	Pred. No.	6.3e-92;				
Matches	557;	Conservative	0;	Mismatches	183;	Indels	12;
						Gaps	1;

175 CGCCCGCTGGACAGGATGTTGCTAGAGGCTGAAGAGGAATATGCTGACCAGGAGA 234

Db 57 CGCGGTCAGGACACAATGTTGCCACGAGCTGAAGAGGAATGTTGGCCACGAGGA 116

235 AGGAGTAGAGGCT-----TTGGCACTGTCCCTTCTATAGCCTGCAGCGACA 282

Db 117 AGACGTGGAGGAGCCCTGGCCGGCTTGAAGACAGTGTCTCTCATACAGCCTGCAGCGCA 176

283 GTCACTCTGGACATGTCCTTGTCAAGCTCCAGCTCTGTACATGCTAGTGAGCCCA 342

Db 177 GTCGCTCCTGGACATGTCCTGGTGAAGTTGCAGCTTGGCCACATGCTTGTGGAGCCCA 236

343 TCTCTGCCGCTCGGTCTCATTCGCCAACACACAGTCCGGCAGATCCAGGAGGAATGAGCCA 402

Db 237 CCTGTGCCGCTCAGTCTCATTTGCCAACACCGTCCGGCAGATCCAAGAGGAGATGACGCA 296

403 GGATGGTGTGGCATGGGATGGCACCAGAAATGTAGATCGGGCACCAGTTGAACGCCCT 462

Db 297 GGATGGGACGTGGCGCACAGTGGCACCACGCTGCAGAGCGGGCGCCGCTCGACCGCTT 356

463 GGTGTCACAGAGATCCTGTGTCGTACACTGAGGGGAGCTGAGGAAGAGCACCCTGCTCC 522

Db 357 GGTCTCCACGAGATCCTGTGCCGTGCAGCGTGGGGGCAAGAGGGGGCACATCCTGCTC 416

523 TGAAGTGGTCTCCCTTGCAAACTCGGTTCCGAGCTCCCATCGTTGGCTCAGC 582

Db 417 TGGCTTGGGGACGCCACACAGGTTCTGACCTTGGCCAGTCACCTCAGC 476

QY	583	CCAGGCGAAGGAACCTCTAGAGAGCCTCTGGGAGATGGAACGCCCAAGAAAACG	64.2
Db	477	ACAGGACCAAGGCACTCTGACAGAGAGCGCTGGGAGATGGATGCGCTTCGAAAAACAG	53.6
QY	643	GGGAAGCTTTGAGAGTCACTGGACCAATATTTTGAGACCCCTGGAGAACAAAACTCCAG	70.2
Db	537	AGGAAGCTTTCACAACTCATTTGATCAGATATTTTGAAACGCTGGAGACTTAAAAACCCAG	58.6
QY	703	TTCAGTGGAGGAACCTCTTCTCAGATGTGGACAGCTCTTACTATGACCTGGACACAGTGCT	76.2
Db	597	CTGCATGGAAGAGCTGTCTCTCAGAGGTGACAGCCCTTACTAGACTGGACACAATGACT	65.6
QY	763	AACAGGAATGATAGTGGAGCCAAAGTCCAGTCTCTGCAATGCGCTTGAGGGCTTTGCTGC	82.2
Db	657	GACAGGACATGATGGGGGGTGCAGGCGCCGGCCCTCTGCAAGGGCTCGAAGGGCTTGGCTCC	71.6
QY	823	AGCCACCCCTCTCTCCAGTTTCCACTTTGCCAAGTCTGACCTGGCTGAGCTGGACCATGTGCT	88.2
Db	717	GGCCACCCCGAGCCCTTAGCTCCAGCTGCAAAATGCCGACTGGGGGAGAGCTGGACCATGCTGT	77.6
QY	883	AGAGATTCTGTGTGAGAGCTCGAAGAGCCACCC	91.4
Db	777	GGAGATCTGTGTGAGAGCTCGAAGAGAGCC	80.8

## RESULT 10

ID AAH06538 standard; cDNA; 862 BP.

AC AAH06538;

DT 26-JUN-2001 (first entry)

DE	Human cDNA clone (5'-primer)	SEQ ID NO:3373.

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; vs.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-00116126.

PR 29-JUL-1999; 99JP-00248036.

11-JAN-2000; 2000JP-00118776.

PR 09-JUN-2000; 2000JP-00241899.

PA (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040

XX 13

length cDNAs defined in the specification, and for the detection and/or

[illegible]

PS Claim 1; SEQ ID NO 3373; 2537pp + Sequence Listing; English.

CC The present invention describes primer sets for synthesizing 5602 full-

CC (a) an oligo-dT primer and an oligonucleotide complementary to the

CC nucleotide sequences defined in the specification, where the

CC of an oligonucleotide comprising a sequence complementary to the

CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
 CC specification. The primer sets can be used in antisense therapy and in  
 CC gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the protein encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH0316 to AAH1362 and  
 CC AAH1363 to AAH1742 represent human cDNA sequences; AAB92446 to AAB95893  
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
 CC oligonucleotides, all of which are used in the exemplification of the  
 CC present invention

Sequence 862 BP; 183 A; 241 C; 286 G; 147 T; 0 U; 5 Other;

Query Match 17.1%; Score 355; DB 4; Length 862;  
 Best Local Similarity 73.7%; Pred. No. 3.2e-75;  
 Matches 484; Conservative 0; Mismatches 160; Indels 13; Gaps 2;

QY 175 CGCCGCTGAGAGATGTTGCTAGAGGGCTGAAGAAATATGTGTACCGAGAGA 234  
 DB 65 CGCGGGTCAGACACAAATGTTGCAAGAGACTGAAGAGAAATGTTGGCCACGAGA 124  
 QY 235 AGAGATAGAGGT-----TTTGGCACTGTCCCTTCTATAGCTGCAGCGACA 282  
 DB 125 AGAGTGAAGGAGCCCTGGCCGGCTTGAAGACAGTGTCTCTATACAGCTTCAGCGGCA 184  
 QY 283 GTCACTCTGAGACATGTGTCCTTTCAGAGCTCCAGCTCTGTCACTGTGTAGAGCCCA 342  
 DB 185 GTCCCTCTGAGACATGTGTCCTTTCAGAGCTCCAGCTCTGTGTAGAGCCCA 244  
 QY 343 TCTTCCGCTGCGTCTCTATCGCCACACAGTCCGAGAGATCCAGAGAAATGAGCCA 402  
 DB 245 CTTGTGCGCTGACGTCTCTATGCGCAACAGTCCGCAATCCAGAGAGATGAGCCA 304  
 QY 403 GGATGCTGTGGCATGGATGGATGCCAACCAGAAATGATCGGCAACCAATTGAAGCCT 462  
 DB 305 GGATGGAGCTGGCGCACAGTGGGACCCAGAGCTGCAGAGCGGGCGCTGCACCGCTT 364  
 QY 463 GGATGCACAGAGATCCTGTGTGTGTACAGTGAAGGAGAGCTGAGAGAAAGACCTTCTCC 522  
 DB 365 GGTCTCCACGAGATCTGT 424  
 QY 523 TGAATGAGATGCTCTCTTGCAGAAATCTGCTTTCAGAGCTCCCATGCTGTGCTCAGC 582  
 DB 425 TTGCTTGGGGGAGCGGCACACACAGGGTCCAGTTTCTGACCTTGGCCAGTCACTCAGC 484  
 QY 583 ACCAGGGGCAAGAGACCTTCAGAGAGCTCTGGAGATGAGACAGCCCAAGAAAAAG 642  
 DB 485 ACAGGAGCAAGAGACCTTCAGAGAGAGCTGGGAGATGATGAGGCTTCGAGAAAAAG 544  
 QY 643 GGGAGCTTTCAGAGATGCTGAGACAGATATTTGAGACCTTGGAGAAAGAAATCTCCAG 702  
 DB 545 AGGAGACTTTCAGAGATGCTGAGATGATGATATTTGAAAGCTGAGAACTTAAACCCAG 604  
 QY 703 TTGAGTGAAGAGAACTTCTTCAGATGAGAGAGCTCTTCTATGAGCTTGAAGAGAGTGC 761  
 DB 605 CTGCACTGGAAGAGCTTCTTCAGAGCTGAGAGAGCTCTTCTATGAGCTTGAAGAGTACT 664  
 QY 762 TAACAGGAATGATGATGAGAGCAAGTCCAGTCTCTGCAATGAGCTTGAAGGCTTTG 818  
 DB 665 TGACAGGCAATGATGAGGAGGTGAGAGGCTCGGAGGCTTCGAGAGGCTTTG 721

RESULT 11  
 AAF94976  
 ID AAF94976 standard; cDNA; 396 BP.  
 AC AAF94976;  
 XX  
 DT 23-MAY-2001 (first entry)

XX  
 DE Human ovarian cancer associated coding sequence SEQ ID NO: 167.  
 XX Human, ovarian cancer; vaccine; gene therapy; carcinoma; ss.  
 XX Homo sapiens.  
 OS  
 PN WO200118046-A2.  
 XX  
 PD 15-MAR-2001.  
 XX  
 PF 08-SEP-2000; 2000WO-US024827.  
 XX  
 PR 10-SEP-1999; 99US-00394374.  
 PR 01-MAY-2000; 2000US-00561778.  
 PR 15-AUG-2000; 2000US-00640173.  
 PR 07-SEP-2000; 2000US-00656668.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Xu J, Stolk JA;  
 XX  
 DR WPI; 2001-211395/21.  
 XX  
 PT Isolated polypeptides associated with ovarian carcinomas, and the nucleic  
 PT acids that encode them, useful for the prevention diagnosis and treatment  
 XX of ovarian cancers.  
 PS Claim 5; Page 170; 189p; English.  
 XX  
 CC The present invention provides a number of coding sequences and proteins,  
 CC the over-expression of which is associated with ovarian carcinoma/cancer.  
 CC These can be used in the diagnosis, treatment and prevention of ovarian  
 CC cancer, optionally by gene therapy or in the form of a vaccine. The  
 CC present sequence is an example of one of these sequences  
 XX  
 SQ Sequence 396 BP; 75 A; 108 C; 147 G; 66 T; 0 U; 0 Other;

Query Match 9.1%; Score 190; DB 4; Length 396;  
 Best Local Similarity 74.7%; Pred. No. 1.3e-35;  
 Matches 257; Conservative 0; Mismatches 75; Indels 12; Gaps 1;

QY 175 CGCCGCTGAGAGATGTTGCTAGAGGGCTGAAGAAATATGTGTACCGAGAGA 234  
 DB 52 CGCGGGTCAGACACAAATGTTGCAAGAGACTGAAGAGAAATGTTGGCCACGAGA 111  
 QY 235 AGAGATAGAGGT-----TTTGGCACTGTCCCTTCTATAGCTTCAGAGCGACA 282  
 DB 112 AGAGTGAAGGAGCCCTGGCCGGCTTGAAGACAGTGTCTCTATACAGCTTCAGAGCGGCA 171  
 QY 283 GTCACTCTGAGACATGTGTCCTTTCAGAGCTTCAGCTCTGTCACTGTGTAGAGCCCA 342  
 DB 172 GTCCCTCTGAGACATGTGTCCTTTCAGAGCTTCAGAGCTTCGTGTAGAGCCCA 221  
 QY 343 TCTTCCGCTGCGTCTCTATCGCCACACAGTCCGAGATCCAGAGAAATGAGCCA 402  
 DB 232 TCTGTGCGGCTCAAGTCTCTATGCGCAACAGTCTCCGCAATCCAGAGAGATGAGCA 281  
 QY 403 GGATGCTGTGGAGATGGATGAGACCCCAAGAAATGATGAGGAGCCAGATTAGCGCTT 462  
 DB 292 GGATGGAGCGTGGGCAACAGTGGACCCCAAGGCTGAGAGAGCGGCGCTCGACCGCTT 351  
 QY 463 GGTGTCCACAGAGATCCTGT 506  
 DB 352 GGTCTCCACGAGATCTGTGCTGTGAGGCTGAGGCTGGGAGCAAGAGG 395

RESULT 12  
 ABL48926  
 ID ABL48926 standard; cDNA; 396 BP.  
 AC ABL48926;  
 XX  
 DT 23-MAY-2001 (first entry)



CC	identifying the expression level in a tissue or cell of at least one gene
CC	listed in the specification. The method is useful for elucidating global
CC	changes in gene expression and for identifying toxicity markers in
CC	tissues or cell exposed to a known toxin. The genes may be used as
CC	toxicity markers in drug screening and toxicity assays. The genes and
CC	gene expression information may be used as diagnostic markers for the
CC	prediction or identification of the physiological state of tissue or cell
CC	sample that has been exposed to a compound or agent. Hepatotoxicity is
CC	characterised by centrilobular necrosis and steatosis. The present
CC	sequence is an expressed sequence tag (EST) or cDNA derived from a gene
CC	which is differentially expressed in response to a hepatotoxic agent
XX	
XX	
SQ	Sequence 522 BP; 149 A; 120 C; 136 G; 117 T; 0 U; 0 Other;
Query Match	8.9%; Score 185.4; DB 6; Length 522;
Best Local Similarity	73.8%; Pred. No. 1.9e-34;
Matches 401; Conservative	0; Mismatches 91; Indels 51; Gaps 11
Oy	1532 GCCTAAGGTTTACTCTCTGGAAATCCCACACACAGTCTTCACAGGACTGTGTGAGGCAGT 159
Dd	519 GTTAAGCGTTCAAGTCGTGAATAATCACAGACACA-TTCTTCAGAAGACTGTGTGACAGGACAGT 461
Oy	1592 GCCATTATGACAGGCTCTGTCTTGTCGGCATCAGTCTGTGGTCCACGCCAGCAGC-AATGGA 165
Dd	460 GCCTTATGACAGGCTGTGCTCTTG-----GGTCCCCAGCCACGACTGTGTGA 414
Oy	1651 CATGAGANATATCAATGCGCCCAACACCAGACACACATGCTCCATGTCOAAAGTGTGACG 1710
Dd	413 CACGAGGACACGACATCCCAACMACAACACAGACACACATGCTATGTCAGAGATATGCG 354
Oy	1711 TGGAGACCACTGGCTCCGCCAGGC---CTGTGCTCAGAGAGGGTGGCAGTCTTAGTG TG 1766
Dd	353 TGGAGACTGACAGGGGTGCCGGGAGTGTGTCTTATGAAAAGGTGTGACGCCACCATGTG 294
Oy	1767 CTGGGGGGGACACACGTGACCTGTGCTTCTTCTTTTAAATAGTGCTTGGACGTTTTA 1826
Dd	293 CT-----CAATPAGAGTACTGCTGCTTGCTGTG-C-TTTAAATAGTGCTTGAACATTTTA 241
Oy	1827 AGGTTAAAAACAAATCCGACTCCATATGATTTTAGGGCTCTCCACCCGCGGGTGGCCCTTA 1886
Dd	240 AGGTTAAAAACAAATGATGTTCTGTA--ACTTAGTGCTCTTCATCTGGGGTGGCCCC-A 184
Oy	1887 TGCTGTCTGCTTGATCTCAAAAGTCTTGCT-----ACTCGGCACTGTCAACTCCA 1937
Dd	183 TGCTGTCTGCTTGGGCTCATAGTCCGATPAACAGATCATGTCAGAACCCACCCCA 124
Oy	1938 CCCCATGATCTCTTTTGTCTCTGTGTGCTTTTTTGGACTTCCCAACTGAGCTTA-- 1995
Dd	123 CCCCAACATGCTTTGTGTTCTCTGTGTGCTTTCTTTTGGACGTCCCACTGAGCCTTAAG 64
Oy	1996 AGGTTTATTTTATATATG-----CTTCATATATCAAATGTAACCTCACTTT 2044
Dd	63 AGGTTTATTTTATATGTTTATGATGATTAGACGTCTTCTTCACTATCAACATGTAAACCTCACTTT 4
Oy	2045 ATT 2047
Dd	3 ATT 1
RESULT 15	
ACH24999	
ID	ACH24999 standard; cDNA; 477 BP.
XX	
AC	ACH24999;
XX	
DT	13-OCT-2003 (first entry)
XX	
DE	Human adult ovary cDNA #3379.
XX	
KW	Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
XX	genome mapping; biodiversity; genetic disorder.
XX	
DS	Homo sapiens.

```
XX US2003073623-A1.
PN
XX
XX 17-APR-2003.
PD
XX
XX 30-JUL-2001; 2001US-00918995.
PF
XX
XX 30-JUL-2001; 2001US-00918995.
PR
XX
XX (DRMA/) DRMANAC R T.
PA (LABA/) LABAT I.
PA (STAC/) STACHE-CRAIN B.
PA (DICK/) DICKSON M C.
XX (JONE/) JONES L W.
XX
XX Dmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
PI
XX
XX WPI; 2003-615964/58.
DR
XX
XX
XX New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX
XX
XX Claim 1; SEQ ID NO 12211; 44pp; English.
PS
XX
XX The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostic as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?docid=20030073623
XX
XX
SQ Sequence 477 BP; 120 A; 78 C; 86 G; 183 T; 0 U; 10 Other;
Query Match 4.7%; Score 98.2; DB 8; Length 477;
Best Local Similarity 78.9%; Pred. No. 1.9e-13;
Matches 157; Conservative 0; Mismatches 33; Indels 9; Gaps 3;
QY 1103 ACTTTCTAAATTAAGTTTATATGTTTGGCAATATTTTGTCTTAAGATATATTTT 1162
Db 61 AATCTTTACAAATTCAGTTTACATG-TTTCGGCAGATTTTGTCTTAAGATATATTTT 119
QY 1163 TAACTTTTATCTTTA-----GATTTTTCAGCTATTTTCTTAAAGTATATTTT 1215
Db 120 TAACTTTTATCTTTAAGTTTATGTTTGGCAATATTTTGTCTTAAGATATATTTT 179
QY 1216 TTCTACAAACATCTCTGCTGTACATTAGAAACATTATTAACCTTAATACGATTGGTGT 1275
Db 180 TTCTATTAACATCTTGTCTGTACATTAG-AACTTTTATAGCTTAACATATGCACTTG 238
QY 1276 GTCATTTTAAAGGTTTAA 1294
Db 239 GTGTGTTTCATTTTAA 257
```

Search completed: September 10, 2004, 11:11:26  
Job time : 769 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 10, 2004, 10:58:30 ; Search time 207 Seconds  
(without alignments)  
5581.680 Million cell updates/sec

Title: US-10-076-069-1

Perfect score: 2082  
Sequence: 1 cccgcgcgtcgtccacc.....atcgaaaaaataaaaaa 2082

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/prodata/2/ina/5A COMB. seq: \*  
2: /cgn2\_6/prodata/2/ina/5B COMB. seq: \*  
3: /cgn2\_6/prodata/2/ina/6A COMB. seq: \*  
4: /cgn2\_6/prodata/2/ina/6B COMB. seq: \*  
5: /cgn2\_6/prodata/2/ina/PTUS COMB. seq: \*  
6: /cgn2\_6/prodata/2/ina/backfile1. seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	425.2	20.4	2629	4 US-09-620-312D-773	Sequence 773, App
2	190	9.1	396	4 US-09-640-173-167	Sequence 167, App
3	190	9.1	396	4 US-09-713-550-167	Sequence 167, App
4	47	2.3	1172	1 US-07-945-288-9	Sequence 9, Appl1
5	47	2.3	1172	1 US-08-462-831-9	Sequence 9, Appl1
6	47	2.3	1172	1 US-08-461-809-9	Sequence 9, Appl1
7	47	2.3	1172	1 US-08-461-441-9	Sequence 9, Appl1
8	47	2.3	1172	5 PCT-US93-08518-9	Sequence 9, Appl1
9	47	2.3	6801	4 US-10-204-708-62	Sequence 62, Appl1
10	47	2.3	19134	2 US-08-487-8268-13	Sequence 13, Appl1
11	47	2.3	640681	4 US-09-780-988-1	Sequence 1, Appl1
12	46.6	2.2	10467	4 US-10-204-708-2	Sequence 2, Appl1
13	46	2.2	665	2 US-08-883-795A-36	Sequence 36, Appl1
14	45.4	2.2	11049	4 US-10-204-708-23	Sequence 23, Appl1
15	45.2	2.2	4140	3 US-08-894-731-2	Sequence 2, Appl1
16	45.2	2.2	6306	4 US-10-204-708-50	Sequence 50, Appl1
17	45.2	2.2	640681	4 US-09-780-988-1	Sequence 1, Appl1
18	45	2.2	6040	4 US-10-204-708-70	Sequence 70, Appl1
19	45	2.2	10619	4 US-10-204-708-3	Sequence 3, Appl1
20	44.8	2.2	658	3 US-08-998-416-595	Sequence 595, App
21	44.6	2.1	2251	3 US-08-991-677-11	Sequence 11, Appl1
22	44.4	2.1	152311	3 US-09-128-155-16	Sequence 16, Appl1
23	44.2	2.1	2447	2 US-09-014-569-14	Sequence 14, Appl1
24	44.2	2.1	5666	4 US-10-204-708-29	Sequence 29, Appl1
25	44.2	2.1	11131	4 US-10-204-708-28	Sequence 28, Appl1
26	43.8	2.1	1447	4 US-09-443-041A-27	Sequence 27, Appl1
27	43.8	2.1	5360	4 US-10-204-708-65	Sequence 65, Appl1

C	28	43.6	2.1	2246	4 US-09-363-708-3	Sequence 3, Appl1
C	29	43.6	2.1	2246	4 US-09-083-587-3	Sequence 3, Appl1
C	30	43.6	2.1	7304	4 US-10-204-708-44	Sequence 44, Appl1
C	31	43.6	2.1	8607	4 US-10-204-708-72	Sequence 72, Appl1
C	32	43.4	2.1	6243	2 US-09-056-075-1	Sequence 1, Appl1
C	33	43.2	2.1	2394	4 US-09-800-729-33	Sequence 33, Appl1
C	34	43.2	2.1	5666	4 US-10-204-708-30	Sequence 30, Appl1
C	35	43.2	2.1	8961	4 US-10-204-708-80	Sequence 80, Appl1
C	36	43	2.1	396	4 US-09-640-173-33	Sequence 33, Appl1
C	37	43	2.1	386	4 US-09-713-550-33	Sequence 33, Appl1
C	38	43	2.1	6866	4 US-10-204-708-20	Sequence 20, Appl1
C	39	42.8	2.1	5219	4 US-10-204-708-51	Sequence 51, Appl1
C	40	42.8	2.1	11049	4 US-10-204-708-22	Sequence 22, Appl1
C	41	42.8	2.1	99916	4 US-09-816-095-3	Sequence 3, Appl1
C	42	42.6	2.0	1390	4 US-09-704-139-1	Sequence 1, Appl1
C	43	42.6	2.0	5562	4 US-10-204-708-63	Sequence 63, Appl1
C	44	42.6	2.0	6409	4 US-09-367-908A-1	Sequence 1, Appl1
C	45	42.4	2.0	4254	2 US-08-443-639-7	Sequence 7, Appl1

## ALIGNMENTS

```
RESULT 1
US-09-620-312D-773
; Sequence 773, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyang
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Wang, Yonghong
; APPLICANT: Wang, Jiah-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungting
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radojelje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_gene Version 1.0
; SEQ ID NO 773
; LENGTH: 2629
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (73)..(798)
; US-09-620-312D-773
```

```
Query Match 20.4%; Score 425.2; DB 4; Length 2629;
Best local similarity 74.1%; Pred. No. 1.4e-109;
Matches 557; Conservative 183; Indels 12; Gaps 1;
QY 175 CGCCGCCCTGGACAGATGTTTGGCTAGAGGCTGAAGAGAAATATGTGACCGAGAGA 234
DB 57 CGCGGTCAGACACATGTTTGCACGAGACTGAAGAGAAATGTGTGGCCACGAGGA 116
```

Qy	235	AGAGGTAGAGGGT-----TTTGGACATGTCCTTCTCTATATAGCTTCGACGCCA	287
Db	117	AGAGCTGAGAGGAGCCCTGGCCGGCTTGAAAGACATGTCTCTCAATACAGCTCGACGGCA	176
Qy	283	GTCATCTCTGGAACATGTCCCTTGTCAAGCTCCAGCTCTGTACATGTGATGGAGCCCA	344
Db	177	GTCGCTCTGGAACATGTCTGTGGAGTTGGACCTTTGGCAACATGTGTGTGGAGCCCA	233
Qy	343	TCCTGCCGCTCGGTCTCTCATTCGCCAACATGTCGGCAGATCCAGAGGAATGAGCCA	402
Db	237	CTGTGCGGCTCAGTCTCTCATTTGCCAACACGGTCCGGCAATCCMAAGAGATGAGCA	296
Qy	403	GGAATGTGTGTGGCATGGATGGCAACCCAGAAATGTATGTGGGCAACAGTTGAAGCCT	466
Db	297	GGATGGGACGTGGGGCAACATGGGACCCAGGCTGCGAGAGCGGGCGCTCGACGGCTT	356
Qy	463	GGTGTCCACAGAAATCTGTGTGTGTACATGTAGAGGGAGCTGAGGAAGAGCACTGTCTC	522
Db	357	GGTCTCCACGGAATCTGTGTCCGTGACGGTGGGGGCAAGAGGGGCAATCTGTCTTC	416
Qy	523	TGAATGAAAGATGTCTCTTGTGAAACTCGGTTTTCCGAGCTCCCATCGTGGCTCAGC	582
Db	417	TGGCTTGGGGGACGGCCACACAGAGGCTCCAGTTTCTGACCTTTGCCAATCACTCAGC	476
Qy	583	ACCAAGGCAAAAGAACCTTCAGACAGACCTCTGGAGATGAGACCCACMAAAAAACAG	642
Db	477	ACAGGCAACMAAGCACTGCAAGACAGCGCTGGAGATGATGGCCCTCGAAMAAAACAG	536
Qy	643	GGGAAGCTTTAGAAATGCACTGGACCAAGATTTTGAACCTCTGGAAACAAAACTCCAG	702
Db	537	AGGAAGCTTTCAACAAATCACTGATATAGATATTGAAACGCTGGAACTTAAAAACCCACG	596
Qy	703	TTCAATGAGGAACCTCTTCTCAGATGTGACAGCTCTACTATGACCTGGACACAGTGT	762
Db	597	CTGCAATGGAAGCTTCTTCTCAGACCTGGAACGCCCTACTAGACTGGACCAATACT	656
Qy	763	AACAGAAATGATGAGTGGGACCAAGTCCAGTCTCTGCAATGAGCTTGAGGGCTTTCTGC	822
Db	657	GACAGGCAATAGGGGGAGTCCAGAGCGGGGCCCTCTGAAAGGCTGAGGGCTTGGCTCC	716
Qy	823	AGCCACCCCTCTCTCCAGTTTCACTTTCAGATCTGACCTGGCTGAGCTGACCATGTGT	882
Db	717	GGCCACCCCAAGCCCTTACCTCCAGTCTGCAAGTCCGACCTGGGAGAGCTGACCACTGCT	776
Qy	883	AGAGATTCTGTGTGAGAACTTGAGAGCCACCC	914
Db	777	GGAGATCTCGTGTGAGAACTTGACAGGAAGCCC	808

```

Result 2
US-09-640-173-167
; Sequence 167, Application US/09640173
; Patent No. 6613515
GENERAL INFORMATION:
APPLICANT: Xu Jiaqun
APPLICANT: Stolk, John A
TITLE OF INVENTION: OVARIAN TUMOR SEQUENCES AND
TITLE OF INVENTION: METHODS OF USE THEREFOR
FILE REFERENCE: 210121.484C2
CURRENT APPLICATION NUMBER: US/09/640.173
CURRENT FILING DATE: 2000-08-15
NUMBER OF SEQ ID NOS: 196
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 167
LENGTH: 396
TYPE: DNA
ORGANISM: Homo sapien
US-09-640-173-167

```

Query Match	9.1%	Score 190;	DB 4;	Length 396;
Best Local Similarity	74.7%;	Pred. No. 1e-43;		
Matches 257; Conservative	0;	Mismatches 75;	Indels 12;	Gaps 1;

QY 175 CGCCCCGCTGACAGAGATGTTTGTCTAGAGAGCTGAAAGAGAAATATGTTGACACAGAGA 234

Db 52 CGCGGGTCAGAGACAATATGTTTGCAAGAGAGCTGAAGAGAAATGTTGTGGCCACAGAGA 111

QY 235 AGAGTAGAGAGGT-----TTTGGCATGTCCCTCTCTATAGCCCTGCAGCGACA 283

Db 112 AGAGCTGGAAGGAGAGCCCTGGCCGGCTTGAAAGACATGTCTCATACAGCCTGCAGCGGCA 171

QY 283 GTCACTCCTTGACATGTCCCTGTGTCAAGCTCCAGCTGTGCATGTGTAAGTGAGAGCCAA 342

Db 172 GTCGCTCCTGGAATGTCTCTGTGAAGTTGAGCTTTGCCACATGCTTTGGAGAGCCAA 231

QY 343 TCTTGTGCCCTCGGTCTCTATGCGCAACAAGTCCGCGAGATCCAGAGAGAAATAGCCA 402

Db 232 TCTGTGCGCTCAAGTCTCATATGGCAACCGGTCCGCGAGATCCAAAGAGAGATAGCA 291

QY 403 GGATGAGTGTGTGGCATGGGATGGGCAACCCCAAGATGTAGTATGGGACACAGTTGAAGCCT 462

Db 292 GGATGGGACCTGTGCGACAGGTGGCAACCCAGAGCTGCAGAGGGGCGCGCTGCACCGCTT 351

QY 463 GGTTGCCACAGAGATCCTGTGTGTAACAAGTAGAGGAGAGCTGAG 506

Db 352 GGTCTCCACGAGATCTTGTGCTGTGACAGCGTGGGGGCAAGAGG 395

```

RESULT 3
US-09-713-550-167
; Sequence 167, Application US/09713550
; Patent No. 6617109
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Scolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.484C4
; CURRENT APPLICATION NUMBER: US/09/713,550
; NUMBER OF SEQ ID NOS: 205
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 167
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-713-550-167

```

Query Match	9.1%;	Score 190;	DB 4;	Length 396;
Best Local Similarity	74.7%;	Pred. No. 1e-43;		
Matches 257;	Conservative	0;	Mismatches 75;	Indels 12;
				Gaps 1;

QY	175	CGCCCGCTGAGACAGGATGTTTGTATAGAGCGCTGAAGAGAAATATGCTGACACAGAAAGA	23
Db	52	CGCGGGTCAGAGACACAATGTTTGCAAGAGAGACTGAAGAGAAATGTTGGCCACAGAGA	111
QY	235	AGAGATGAGAGGT-----TTTGGCACTGTCCCTTCTTAATAGCTTGCAGCGACA	287
Db	112	AGAGGTGAGAGGAGCCCTGAGCGGCTTGTAAGACAATGTCTCATACACTTGACGGGCA	171
QY	283	GTCATCTCTGGAATGTCCCTTGTAAGCTCAGCTCTGTACAATGCTAATGAGAGCCCA	342
Db	172	GTCGCTCTTGAATATCTCTGTGTAAGTTGACAGCTTTGCCACATGCTTGTGGAGCCAA	2311
QY	343	TCTCTGCGCTCGGTCTTCATCGCCACAACAGTCCGGAGATCCAGAGAGAAATGAGCCA	402
Db	232	TCTGTGCGCTCAAGTCTTCATTGCCAACACGGTCCGGCAGATCCAAAGAGAGATGACGA	2911
QY	403	GGATGAGTGTGGCATGGAGTGGACCCACAGATATGTATCATCGGGACACAGTTAAAGCCT	462
Db	292	GGATGGGACGTGGCGCACAGTGGACACCCAGGCTGTCAAGCGGAGCGCCGCTCTACCGCTT	351
QY	463	GGTGTCCACAGAGATCTGTGTCTGATCAAGTGAAGGGAGAGCTGAGG	506
Db	352	GGTCTCCACGAGATCTGTGTCCGTGGACGCGTGGGGGACAGAGG	395



```

APPLICANT: T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
TITLE OF INVENTION: DERMATOPHAGOCIDES
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 STATE STREET, SUITE 510
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,831
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/945,288
FILING DATE: 10 SEPTEMBER 1992
APPLICATION NUMBER: US 580,655
FILING DATE: 11 SEPTEMBER 1990
APPLICATION NUMBER: US 458,642
FILING DATE: 13 FEBRUARY 1990
ATTORNEY/AGENT INFORMATION:
NAME: MANDRAGONURAS, AMY E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1172 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..738
US-08-462-831-9

Query Match      2.3% Score 47; DB 1; Length 1172;
Best Local Similarity 56.0%; Pred. No. 0.005;
Matches    89; Conservative   0; Mismatches    70; Indels     0; Gaps    0;

Qy      1119 TTATAAATGTTTTGGCAAAATTTGTCTTAGATATATTTTAAACTTTATACCT 1178
          |||||            |         |||||             |||||
Db       1169 TTTTTCCTTTTCCTTTTCCTTTTCCTTTTCCTTTTCCTTTTCCTTTTCCT 1110
               |||||            |         |||||             |||||

Qy      1179 TAGATTTTTCAGCTATTTTCCTTAAGAATATTTTCTACAACATCTCGTGCT 1238
          |||||            |         |||||             |||||
Db       1109 TTTTTTTTAAAAATTTTGTGAATGAATTTATTCATTTCTAAAATAATTTGATTTTAAAT 1050
                |||||            |         |||||             |||||

Qy      1239 ACATTGAACATTATTAACCTAAATACATGTGCT 1277
          |||||            |         |||||             |||||
Db       1049 AAATTAGTGAACATCATATAGAAATTTGCTTTTGT 1011
                |||||            |         |||||             |||||


RESULT 6
US-08-461-809-9/c
Sequence 9, Application US/08461809
Patent No. 5770202
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
TITLE OF INVENTION: DERMATOPHAGOCIDES
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
```

```

ADDRESS: LAHIVE & COCKFIELD
STREET: 60 STATE STREET, SUITE 510
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,809
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/945,288
FILING DATE: 10 SEPTEMBER 1992
APPLICATION NUMBER: US 580,655
FILING DATE: 11 SEPTEMBER 1990
APPLICATION NUMBER: US 458,642
FILING DATE: 13 FEBRUARY 1990
ATTORNEY/AGENT INFORMATION:
NAME: MANDRAGOURAS, AMY E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1172 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..738
US-08-461-809-9

Query Match
Best Local Similarity 2.3%; Score 47; DB 1; Length 1172;
Matches 89; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 1119 TTTTATGTTTTGGCAATATTTGCTTAGATATATTTTAACTTTTAACTT 1178
DB 1169 TTTTATGTTTTGGCAATATTTGCTTAGATATATTTTAACTTTTAACTT 1110
QY 1179 TAGATTTTTCAGCTATTTTCTTAAAGTATATTTTCTACAAACATCTCTGCTGCT 1238
DB 1109 TTTTATGTTTTGGCAATATTTGCTTAGATATATTTTAACTTTTAACTT 1050
QY 1239 ACATTAGAACATTTTAACTTAAATAGATGATGCTGT 1277
DB 1049 AATTAGTACATCATTAATAGAAATGCTTTTGT 1011

RESULT 7
US-08-461-441-9/c
Sequence 9, Application US/08461441
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESS: LAHIVE & COCKFIELD
STREET: 60 STATE STREET, SUITE 510
CITY: BOSTON
STATE: MA
COUNTRY: USA
```

```

ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,441
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/945,288
FILING DATE: 10 SEPTEMBER 1992
APPLICATION NUMBER: US 580,655
FILING DATE: 11 SEPTEMBER 1990
APPLICATION NUMBER: US 458,642
FILING DATE: 13 FEBRUARY 1990
ATTORNEY/AGENT INFORMATION:
NAME: MANDRAGOURAS, AMY E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1172 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..738
US-08-461-441-9

Query Match
Best Local Similarity 2.3%; Score 47; DB 1; Length 1172;
Matches 89; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 1119 TTTTATGTTTTGGCAATATTTGCTTAGATATATTTTAACTTTTAACTT 1178
DB 1169 TTTTATGTTTTGGCAATATTTGCTTAGATATATTTTAACTTTTAACTT 1110
QY 1179 TAGATTTTTCAGCTATTTTCTTAAAGTATATTTTCTACAAACATCTCTGCTGCT 1238
DB 1109 TTTTATGTTTTGGCAATATTTGCTTAGATATATTTTAACTTTTAACTT 1050
QY 1239 ACATTAGAACATTTTAACTTAAATAGATGATGCTGT 1277
DB 1049 AATTAGTACATCATTAATAGAAATGCTTTTGT 1011

RESULT 8
PCT-US93-08518-9/c
Sequence 9, Application PC/TUS9308518
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESS: LAHIVE & COCKFIELD
STREET: 60 STATE STREET, SUITE 510
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/08518
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/945,288
; FILING DATE: 10 SEPTEMBER 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: MANDRAGOURAS, AMY E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1172 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..738
; PCT-US93-08518-9

Query Match      2.3%; Score 47; DB 5; Length 1172;
Best Local Similarity 56.0%; Pred. No. 0.005;
Matches 89; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 1119 TTTTATGTTTGGGCAATTTTGTCTAGATATTTTAACTTTTAACTT 1178
DB 1169 TTTTATGTTTGGGCAATTTTGTCTAGATATTTTAACTTTTAACTT 1110

QY 1179 TAGATTTTTCAGCTATTTCTTAAAGATATTTTCTACAAATCTCTGCTGCT 1238
DB 1109 TTTTATGTTTGAATGATGATTTATTCATTTTAAATAATTTGATTTAAAT 1050

QY 1239 ACATTGAAACATTTTATACCTTAATACGTTGCTGCT 1277
DB 1049 AATTAGTGACATCATATAGAAATTTCTTTTGT 1011

RESULT 9
US-10-204-708-62
; Sequence 62, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; PRIOR FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 62
; LENGTH: 6801
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
```

```

US-10-204-708-62

Query Match      2.3%; Score 47; DB 4; Length 6801;
Best Local Similarity 51.2%; Pred. No. 0.016;
Matches 110; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 1103 ACTTCTAAATTAAGTTTATATGTTTGGGCAATTTTGTCTAGATATTTT 1162
DB 5570 ATTTATATTTTAGTAAAGTTTCTTTTGTAAATTTAAATTTTAAATTTT 5629

QY 1163 TAACTTTTATACCTTATGATTTTTCAGCTATTTCTTAAAGATATTTTCTACA 1222
DB 5630 ATTTAGTTTCTTTTATATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 5689

QY 1223 AACATCTCTGCTGCTACATTTAGAAACATTTTAACTTAATACGATTTGTCATTT 1282
DB 5690 TTTTATTTTGGTTTATTTTAAATTTTCTTTTAAATTTGATATTCGTTTGTATTT 5749

QY 1283 TAAAGTTTAAATAGAAACCTTTGTTTACTGA 1317
DB 5750 AATTTTCTTTGTTAAAGTTATTTTATTTATTTAA 5784

RESULT 10
US-08-487-826B-13/C
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelisen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH21.001CPI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-487-826B-13

Query Match      2.3%; Score 47; DB 2; Length 19124;
Best Local Similarity 51.2%; Pred. No. 0.031;
```

RESULT 12  
US-10-204-708-2  
Sequence 2, Application US/10204708  
Patent No. 6677731  
GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: PIEPENBROCK, Christian  
APPLICANT: BERLIN, Kurt  
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication  
TITLE OF INVENTION: by Assessing DNA Methylation

RESULT 13  
 US-08-883-795A-36  
 ; Sequence 36, Application US/08883795A  
 ; Patent No. 5985607  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Delcive, Genevieve  
 ; APPLICANT: Awang, Gregor  
 ; TITLE OF INVENTION: Recombinant DNA Molecules and Expression  
 ; TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator  
 ; NUMBER OF SEQUENCES: 39  
 ; CORRESPONDENCE ADDRESSES:  
 ; ADDRESSEE: BERESKIN & PARR  
 ; STREET: 40 King Street West  
 ; CITY: Toronto  
 ; STATE: Ontario  
 ; COUNTRY: Canada  
 ; ZIP: M5H 3Y2  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/883,795A  
 ; FILING DATE: 27-JUN-1997  
 ; CLASSIFICATION: 435

```

;
; ATTORNEY/AGENT INFORMATION:
; NAME: Gravelle, Michelle
; REGISTRATION NUMBER: 40,761
; REFERENCE/DOCKET NUMBER: 7841-062
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 364-7311
; TELEFAX: (416) 361-1398
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 665 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: Rh 32
; US-08-883-795A-36

Query Match          2.2%; Score 46; DB 2; Length 665;
Best Local Similarity 52.0%; Pred. No. 0.0067;
Matches 189; Conservative 0; Mismatches 115; Indels 4; Gaps 1;

QY 1066 ACAGCTATGCTGCTATTTTCATTTAGATGAGTGAACCTTCTAAATAATTAAGTTTATA 1125
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 50 ATATTTATATTAATAATTTATTAATTTATTAATTTATTAATTTATTAATTTATTAATTA 109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1126 TGTTTTGGCAATATTTTGTCTTAAGATATATTTT---TTAACTTTTATACCTTTAG 1181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 110 TATTTATATATTAATAATTTATTAATTTATTAATTTATTAATTTATTAATTTATTA 169
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1182 ATTTTTCAGCTATTTTCTTAAGAATTTTTCACAAACATCCTGCTGTACA 1241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 170 TATTTATATTAATAATTTATTAATTTATTAATTTATTAATTTATTAATTTATTA 229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1242 TTGAACAATTTTAACCTTAATACGATTGCTGTCATTTTAAGTTTAATTAAGAAA 1301
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 230 TATTTATATTAATAATTTATTAATTTATTAATTTATTAATTTATTAATTTATTA 289
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1302 CTTCCTTT 1309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 290 TATTTTAT 297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
US-10-204-708-23
; Sequence 23, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10035229.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 23
; LENGTH: 11049
; TYPE: DNA
; ORGANISM: Artificial Sequence
```

```

;
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; US-10-204-708-23

Query Match          2.2%; Score 45.4; DB 4; Length 11049;
Best Local Similarity 52.3%; Pred. No. 0.061;
Matches 124; Conservative 0; Mismatches 111; Indels 2; Gaps 1;

QY 1112 AATTAAGTTTATATGCTTTTGGCAATATTTGCTTAAAGATATTTTAAACTTT 1171
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8235 ATTTAGTTTTTTTTTATTTATTTAGACGTTTTTTTTTTTTTTTATTAAGATTTT 8294
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1172 TATACCTTAGATTTTTCAGCTATTTTCTTAAGATATTTTCTCAAAACATCTTC 1231
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8295 TTAATTTTATTTTATTTTATTTGTTTATATTAATTAATTTATTTTGAATTTA 8354
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1232 TGTCTACATTAGAAACAT--TTATTAACCTTAATACGATGCTGTCATTTTAAAGT 1289
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8355 TTTTATAGAAATAGAAATAATGTTATTTTATATTTTGGTTGAATGTTATTTT 8414
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1290 TTAATAGAAACTCTTTTGTACGATCTCTACATCCCAAGCACTGTAAAT 1346
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8415 TTAATAAGAAATTTTATGCTATTAATTAAGAAAAATGAATTTGTTAAT 8471
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
US-08-894-731-2
; Sequence 2, Application US/08894731
; Patent No. 6084089
; GENERAL INFORMATION:
; APPLICANT: MINE, Toshiki
; APPLICANT: OHYAMA, Akio
; APPLICANT: HIYOSHI, Toru
; APPLICANT: KASAKA, Keisuke
; TITLE OF INVENTION: COLD-INDUCIBLE PROMOTER SEQUENCE
; FILE REFERENCE: 760-234P
; CURRENT APPLICATION NUMBER: US/08/894,731
; CURRENT FILING DATE: 1997-10-27
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 4140
; TYPE: DNA
; ORGANISM: Solanum tuberosum
; US-08-894-731-2

Query Match          2.2%; Score 45.2; DB 3; Length 4140;
Best Local Similarity 51.5%; Pred. No. 0.037;
Matches 104; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 1122 TATATGTTTTGGCAATATTTTGTCTTAAGATATTTTAACTTTTATACCTTTAG 1181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 237 TATATATTTATTAATAATAATTAAGCTATTAATTAAGATATTTTATTTTACCTTTT 296
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1182 ATTTTTCAGCTATTTTCTTAAGATATTTTCTCAAAACATCCTGCTGTACA 1241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 297 AGTTATTTTCAAAATATTTTAAAAATATTAATTAATTAATTAATTAATTAATTA 356
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1242 TTGAACAATTTTAACCTTAATACGATTGCTGTCATTTTAAGTTTAATTAAGAAA 1301
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 357 TATTAATATCTCTGAATTTATTAAGTAAATAATTAATTAATTAATTAATTAATTA 416
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1302 CTTCCTTTTGTACGATCTCT 1323
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 417 ATTAATTAATTAATTAATCTCT 438
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Search completed: September 10, 2004, 14:20:52
Job time : 210 secs
```

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 10, 2004, 12:42:41 : Search time 699 Seconds  
(without alignments)  
14986.556 Million cell updates/sec

Title: US-10-076-069-1

Perfect score: 2082

Sequence: 1 ccccgctgcgtctccacc.....atcgaaaaaaaaaaaaa 2082

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3304383 seqs, 2515761380 residues

Total number of hits satisfying chosen parameters: 6608766

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2\_6/prodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/prodata/1/pubpna/PCR\_NEW\_PUB.seq:\*  
3: /cgn2\_6/prodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/prodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/prodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/prodata/1/pubpna/PCRUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/prodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/prodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/prodata/1/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/prodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/prodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/prodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/prodata/1/pubpna/US09\_NEW\_PUB.seq2:\*  
14: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq:\*  
15: /cgn2\_6/prodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
16: /cgn2\_6/prodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
17: /cgn2\_6/prodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
18: /cgn2\_6/prodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
19: /cgn2\_6/prodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2082	100.0	2082	14 US-10-076-069-1	Sequence 1, Appl1
2	428.4	20.6	2038	9 US-09-925-302-153	Sequence 153, App
3	428.4	20.6	2038	13 US-09-925-302-153	Sequence 153, App
4	428.4	20.6	2038	15 US-09-925-302-153	Sequence 153, App
5	426.8	20.5	2082	14 US-10-076-069-1	Sequence 179, Appl
6	425.2	20.4	2447	13 US-10-362-893-5	Sequence 3, Appl1
7	425.2	20.4	2629	15 US-10-037-270-773	Sequence 5, Appl1
8	425.2	20.4	2629	16 US-10-117-722-773	Sequence 773, App
9	190	9.1	386	9 US-09-825-294-167	Sequence 167, App
10	190	9.1	386	9 US-09-970-966-167	Sequence 167, App
11	190	9.1	386	15 US-10-212-677-167	Sequence 167, App
12	190	9.1	386	15 US-10-361-811-167	Sequence 167, App
13	190	9.1	386	16 US-10-369-186-167	Sequence 167, App
14	185.4	8.9	522	9 US-09-917-800A-1285	Sequence 1285, Ap

15	98.2	4.7	477	10 US-09-918-995-12211	Sequence 12211, A
16	62.4	3.0	368	9 US-09-867-701-3214	Sequence 3214, Ap
17	61.2	2.9	18988	13 US-10-221-714A-64	Sequence 64, Appl
18	61.2	2.9	18988	15 US-10-311-455-674	Sequence 674, Appl
19	61.2	2.9	18988	15 US-10-240-485-62	Sequence 62, Appl
20	60.4	2.9	3673778	15 US-10-312-841-1	Sequence 1, Appl1
21	59	2.8	17721	15 US-10-311-455-1701	Sequence 1701, Ap
22	56.2	2.7	9642	15 US-10-311-455-329	Sequence 329, App
23	55.8	2.7	960	15 US-10-198-846-6381	Sequence 6381, Ap
24	55.2	2.7	5689	13 US-10-221-714A-148	Sequence 148, App
25	55.2	2.7	5689	15 US-10-239-676-90	Sequence 90, Appl
26	55.2	2.7	5689	15 US-10-240-453-100	Sequence 100, App
27	55.2	2.7	83391	17 US-10-433-793-123	Sequence 123, App
28	54.8	2.6	15548	15 US-10-311-455-2128	Sequence 2128, App
29	54.6	2.6	5488	15 US-10-311-455-1429	Sequence 1429, Ap
30	54.4	2.6	33053	17 US-10-433-793-36	Sequence 36, Appl
31	54.4	2.6	37973	15 US-10-311-455-2169	Sequence 2169, Ap
32	54	2.6	8136	17 US-10-311-455-528	Sequence 528, App
33	54	2.6	8136	17 US-10-257-166-38	Sequence 38, Appl
34	53.8	2.6	7589	15 US-10-240-453-264	Sequence 264, App
35	53.6	2.6	113515	15 US-10-311-455-2147	Sequence 2147, Ap
36	53.4	2.6	6128	15 US-10-311-455-29	Sequence 29, Appl
37	52.6	2.5	6904	15 US-10-311-455-185	Sequence 185, Appl
38	52.2	2.5	5273	15 US-10-311-455-848	Sequence 848, App
39	52	2.5	15161	13 US-10-221-613-385	Sequence 385, App
40	52	2.5	3673778	15 US-10-312-841-2	Sequence 2, Appl1
41	51.8	2.5	591	17 US-10-437-963-21828	Sequence 21828, A
42	51.8	2.5	6154	13 US-10-221-613-70	Sequence 70, Appl
43	51.6	2.5	17538	15 US-10-311-455-1130	Sequence 1130, Ap
44	51.6	2.5	516	9 US-09-960-352-5785	Sequence 5785, Ap
45	51.6	2.5	6478	15 US-10-239-676-124	Sequence 124, App

## ALIGNMENTS

RESULT 1  
US-10-076-069-1  
; Sequence 1, Application US/10076069  
; Publication No. US20020177214A1  
; GENERAL INFORMATION:  
; APPLICANT: JURECIC, ROLAND  
; APPLICANT: NACHTMAN, RONALD  
; TITLE OF INVENTION: HELP, A NOVEL GENE WITH A ROLE IN HEMATOPOIETIC AND NEURAL DEVEL  
; FILE REFERENCE: 39532-176599  
; CURRENT APPLICATION NUMBER: US/10/076, 069  
; CURRENT FILING DATE: 2002-02-15  
; PRIOR APPLICATION NUMBER: US 60/268, 923  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 2082  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (191)..(901)  
US-10-076-069-1

Query Match 100.0%; Score 2082; DB 14; Length 2082;  
Best Local Similarity 100.0%; Pred No. 0;  
Matches 2082; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ccccgctgcgtctccacccttgcagctgcgcgcgcttgcgcagtttc 60  
1 ccccgctgcgtctccacccttgcagctgcgcgcgcttgcgcagtttc 60  
Db 1 ccccgctgcgtctccacccttgcagctgcgcgcgcttgcgcagtttc 60  
QY 61 gggggactggaacttcccttgcagctgcgcgcgcttgcgcagtttc 120  
61 gggggactggaacttcccttgcagctgcgcgcgcttgcgcagtttc 120  
Db 61 gggggactggaacttcccttgcagctgcgcgcgcttgcgcagtttc 120  
121 gcaagtgacgagttccggggtcccgaaagagctggggtgctgcttgcagcgcgcgcg 180

```

Db      121  GCAGGTGACCGATTCCGGGTGCCGAAAGAGCTGGCGTGTCTGCTTGCAGCCGCCG 180
Qy      181  CCTGGACAGGATGTTTGTCTAAGAGGCTGAAGAGAAATATGTGTGACCAAGAAAGAGT 240
Db      181  CCTGGACAGGATGTTTGTCTAAGAGGCTGAAGAGAAATATGTGTGACCAAGAAAGAGT 240
Qy      241  AGAGGGTTTTGGSCATGTCCCTTCTATAGCTGCAGAGACATCTCCCTGACATGTCTC 300
Db      241  AGAGGGTTTTGGSCATGTCCCTTCTATAGCTGCAGAGACATCTCCCTGACATGTCTC 300
Qy      301  CTTTGTCAAGCTCCAGCTCTGTCAATGTCTAGTGAAGCCCAATCTCTGCGCTGTCTCT 360
Db      301  CTTTGTCAAGCTCCAGCTCTGTCAATGTCTAGTGAAGCCCAATCTCTGCGCTGTCTCT 360
Qy      361  CATGCCCAACACAGTCCGGCAGATCCAGAGAGAAATAGACCAAGATGTGTGTGGATGG 420
Db      361  CATGCCCAACACAGTCCGGCAGATCCAGAGAGAAATAGACCAAGATGTGTGTGGATGG 420
Qy      421  GATGGACACCCAGAAATGTAGATGGGGACCAAGTTGAACGCTGTGTCTCAAGAGATCCT 480
Db      421  GATGGACACCCAGAAATGTAGATGGGGACCAAGTTGAACGCTGTGTCTCAAGAGATCCT 480
Qy      481  GTGTGTACAGTGAAGGGAGCTGAAGAAAGACACCTGTCTCTGAACCTGGAAATGTCTCC 540
Db      481  GTGTGTACAGTGAAGGGAGCTGAAGAAAGACACCTGTCTCTGAACCTGGAAATGTCTCC 540
Qy      541  CTTGGCAAACTCGGTTTTCCGAGTCCCAATCTGTGTGTGTGTGTGTGTGTGTGTGTGT 600
Db      541  CTTGGCAAACTCGGTTTTCCGAGTCCCAATCTGTGTGTGTGTGTGTGTGTGTGTGTGT 600
Qy      601  TCAGAGAGGCTCTGGAGATGAGACAGCCCAAGAAACAGGGGAAAGCTTTGAGAAATGC 660
Db      601  TCAGAGAGGCTCTGGAGATGAGACAGCCCAAGAAACAGGGGAAAGCTTTGAGAAATGC 660
Qy      661  ACTGAGACCAATATTTTGAACCTCTGAGAAACAAATCTCCAGTTCACTGTGAGAGACTCT 720
Db      661  ACTGAGACCAATATTTTGAACCTCTGAGAAACAAATCTCCAGTTCACTGTGAGAGACTCT 720
Qy      721  CTGAGATGTGAGACAGCTCCCTAATATGACTGTGACACAGTGTCTTAAACAGAAATGTAG 780
Db      721  CTGAGATGTGAGACAGCTCCCTAATATGACTGTGACACAGTGTCTTAAACAGAAATGTAG 780
Qy      781  GACCAAGTCCAGTCTGTCAATGAGGCTTGAAGGCTTGTGTGAGCCACCCCTCTCCAG 840
Db      781  GACCAAGTCCAGTCTGTCAATGAGGCTTGAAGGCTTGTGTGAGCCACCCCTCTCCAG 840
Qy      841  TTTCACCTTGAAGTCTGACCTGGCTGAGCTGACACATGTGTGAGAGATTTCTGTGAGAC 900
Db      841  TTTCACCTTGAAGTCTGACCTGGCTGAGCTGACACATGTGTGAGAGATTTCTGTGAGAC 900
Qy      901  CTGAGAGGACACCCAGTGGGCTAAGGGTGAAGGCCACAGTCCCATGAGCTCAAGTGT 960
Db      901  CTGAGAGGACACCCAGTGGGCTAAGGGTGAAGGCCACAGTCCCATGAGCTCAAGTGT 960
Qy      961  GTTGTGACCCAGACAGATTAAGCATTTGTCTTAAGAGGGGCTGTGGCTTTGAGCTCAT 1020
Db      961  GTTGTGACCCAGACAGATTAAGCATTTGTCTTAAGAGGGGCTGTGGCTTTGAGCTCAT 1020
Qy      1021  TATCCTTTTGTGTGACATTTGAGCTCACTGTGAGAGATGTGTGTCTGACACTATGTCTAGT 1080
Db      1021  TATCCTTTTGTGTGACATTTGAGCTCACTGTGAGAGATGTGTGTCTGACACTATGTCTAGT 1080
Qy      1081  CTATTTTCAATTAAGTAGGTAAGCTTCTTAATTAAGTTTAAATGTTTAAATGTTTGGCAATA 1140
Db      1081  CTATTTTCAATTAAGTAGGTAAGCTTCTTAATTAAGTTTAAATGTTTAAATGTTTGGCAATA 1140
Qy      1141  TTTTGTCTTAAGATATATTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTT 1200
Db      1141  TTTTGTCTTAAGATATATTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTTAACTTTT 1200
Qy      1201  TTTAAAGTATATTTTCTTACAAACATCTGTGTGTCTTACATTTAAGAACTTTATTAACCT 1260
Db      1201  TTTAAAGTATATTTTCTTACAAACATCTGTGTGTCTTACATTTAAGAACTTTATTAACCT 1260

Db      1201  TTTAAAGTATATTTTCTTACAAACATCTGTGTGTCTTACATTTAAGAACTTTATTAACCT 1260
Qy      1261  AATTAAGATGTGTGTCTATTTTAAAGTTTAAATAGAAATCTTCTTTGTACTAGATCTC 1320
Db      1261  AATTAAGATGTGTGTGTCTATTTTAAAGTTTAAATAGAAATCTTCTTTGTACTAGATCTC 1320
Qy      1321  TCTAAGCTCCCAAGGCACTGTAAATGTAGCCGGCCGGGTGTCTTACATGAGAGCTCCAG 1380
Db      1321  TCTAAGCTCCCAAGGCACTGTAAATGTAGCCGGCCGGGTGTCTTACATGAGAGCTCCAG 1380
Qy      1381  TATGCTCTATCTTATGTAGAGCTTGAAGAAACATGACAGCTCCATGCCCCCTCAC 1440
Db      1381  TATGCTCTATCTTATGTAGAGCTTGAAGAAACATGACAGCTCCATGCCCCCTCAC 1440
Qy      1441  TGGGTCTGTCTGGCGGATCGAGAGCTCTTCTTCTAGCCCGGTGTGACAGATGGCTTTAT 1500
Db      1441  TGGGTCTGTCTGGCGGATCGAGAGCTCTTCTTCTAGCCCGGTGTGACAGATGGCTTTAT 1500
Qy      1501  TATGCTATTTATATGTAAATGCACTGAAAGCTTAAGTCTTATCTCTGAAATCCCAAC 1560
Db      1501  TATGCTATTTATATGTAAATGCACTGAAAGCTTAAAGTCTTATCTCTGAAATCCCAAC 1560
Qy      1561  ACCAGTCTTCAAGGACCTGTGAGGCGAGTCCCTTATGACAGGCTTGTCTTGGCCATC 1620
Db      1561  ACCAGTCTTCAAGGACCTGTGTGAGGCGAGTCCCTTATGACAGGCTTGTCTTGGCCATC 1620
Qy      1621  ACTGTCTGTTCCAGCCAGCAGCATGTGACATGAGCATGACATGACATGACATGACATGAC 1680
Db      1621  ACTGTCTGTTCCAGCCAGCAGCATGTGACATGAGCATGACATGACATGACATGACATGAC 1680
Qy      1681  GCACCAATGCTCCCAATGCAAGTGTGACATGAGAGACCACTGTGTCTCCAGGCTGTGTCT 1740
Db      1681  GCACCAATGCTCCCAATGCAAGTGTGACATGAGAGACCACTGTGTCTCCAGGCTGTGTCT 1740
Qy      1741  AGAGAGGCTGTGCAATCTTCAAGTGTGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
Db      1741  AGAGAGGCTGTGCAATCTTCAAGTGTGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
Qy      1801  TTTTAAATGCTGTGAGCTTTTAAAGTTTAAATCAATCCGACTCCATATGATTTAGG 1860
Db      1801  TTTTAAATGCTGTGAGCTTTTAAAGTTTAAATCAATCCGACTCCATATGATTTAGG 1860
Qy      1861  GCTCTCCACCTGGGGGTGGCCCTAATGCTGTGTGTGATCTTCAAAAGTCTTGTGATCTC 1920
Db      1861  GCTCTCCACCTGGGGGTGGCCCTAATGCTGTGTGTGATCTTCAAAAGTCTTGTGATCTC 1920
Qy      1921  GGCACGTGCAAGCTCCACCCCATGTATCTTTTGTCTGTGTGTGTGTGTGTGTGTGTGTGT 1980
Db      1921  GGCACGTGCAAGCTCCACCCCATGTATCTTTTGTCTGTGTGTGTGTGTGTGTGTGTGTGT 1980
Qy      1981  CCCAAGCTGAGCTTAAGTTTATTTTATATGTGCTTCAATATCAACATGTAAACCTCA 2040
Db      1981  CCCAAGCTGAGCTTAAGTTTATTTTATATGTGCTTCAATATCAACATGTAAACCTCA 2040
Qy      2041  CTTTATTTAAAGTATCCAGCAATATGAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAA 2082
Db      2041  CTTTATTTAAAGTATCCAGCAATATGAAAAAATGAAAAAATGAAAAAATGAAAAAATGAAAA 2082

RESULT 2
US-09-925-302-153
; Sequence 153. Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
```



```
/ NUMBER OF SEQ ID NOS: 896
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 153
/ LENGTH: 2038
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (1490)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: misc feature
/ LOCATION: (1508)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: misc feature
/ LOCATION: (1979)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: misc feature
/ LOCATION: (1992)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: misc feature
/ LOCATION: (2010)
/ OTHER INFORMATION: n equals a,t,g, or c
/ US-09-925-302-153
```

```
Query Match      20.6%; Score 428.4; DB 9; Length 2038;
Best Local Similarity 74.3%; Pred. No. 6.9e-98;
Matches 559; Conservative 0; Mismatches 181; Indels 12; Gaps 1;
```

```
QY 175 CGCCGCCCTGGACAGATGTTTCTAGAGGGCTGAAGAGAAATATGTGACCGAGAGA 234
    |||||
DB 82 CGGGGGTCAGAGACACAATGTTTGCAGAGACTGAAGAGAAATGTTGGCCACGAGGA 141
    |||||
QY 235 AGAGATGAGAGGT-----TTGGCACTGTCCTTCTTATGCTTGCAGGACA 282
    |||||
DB 142 AGACGTGAGGGGGCCCTGGCCGCTTGAAGACAGTCTCTCATACGCTTGCAGGGCA 201
    |||||
QY 283 GTCACTCTGGAGATGTCCTTGTCAAGCTCCAGCTGTGCATGCTAGTGGAGCCCA 342
    |||||
DB 202 GTGCTCTGACATGTCCTTGTGTAAGTTGACGCTTTCACATGCTTGTGAGCCCA 261
    |||||
QY 343 TCTTCCCGCTCGGTCTCATTCGCCAACACAGTCCGGCAGATCCAGAGAGAAATGAGCA 402
    |||||
DB 262 TCTGTGCGGCTCAGTCTCTCATTCGCCAACACAGTCCGGCAGATCCAGAGAGAAATGAGCA 321
    |||||
QY 403 GGAATGTTGTGGATGGAGTGGACATCCCGAATGTAATGCGGCACAGTGAAGCCCT 462
    |||||
DB 322 GGAATGGAAGTGGCGGACAGTGGACATCCCGCAGCTGCGAGACGGGCGCTTGCACGCTT 381
    |||||
QY 463 GGTGTCCACAGATCTGTGTGTACAGTGAAGGGGAGCTGAAGAAAGAGACCCCTGCTCC 522
    |||||
DB 382 GGTGTCCACAGATCTGTGTGTACAGTGAAGGGGAGCTGAAGAAAGAGACCCCTGCTCC 441
    |||||
QY 523 TGAATGGAAGATGCTCTTGTGCAAAATCGGTTTCGAGCTCCCAATCGTTGCTCAGC 582
    |||||
DB 442 TGTCTTGGGGGAGGGGACACACAGGGTCCAGTTTTCAGCTTTGGCCCATCCTCAGC 501
    |||||
QY 583 ACCAGGGCAAGAGACCTCTCAGAGCAGCTCTGGAGATGACAGCCACAGAAAACAG 642
    |||||
DB 502 ACAGGACACCAAGGACCTGCGAGAGCAGCGCTGGAGATGATGCGCTCGAAGAAAACAG 561
    |||||
QY 643 GGAAGCTTTCAGAGTCACTGACACAGATATTGAGACCTGGAGAACAAAATCTCCAG 702
    |||||
DB 562 AGGAAGCTTTCAGAGTCACTGATATGATATTTGAAACGCTGAGATTAAGAAAACCCAG 621
    |||||
QY 703 TTCAATGGAAGATCTTCTCAGATGTGACAGCTCTCACTATGACCTGACACAGTCT 762
    |||||
DB 622 CTGCAATGGAAGATCTTCTCAGATGTGACAGCTCTCACTATGACCTGACACAGTCT 681
    |||||
QY 763 AACAGATATGATGATGAGGACCAAGTTCAGTCTCTGCAATGAGCTTGAAGGCTTTGCTGC 822
    |||||
DB 682 GACAGGATGATGATGAGGAGTCCAGGCGGGGCGCTTGGAGAGGCTCGAGGGCTTGGCTCC 741
    |||||
QY 823 AGCAGCCCTCTCTCCAGTTCCAGTTCAGATGCTGACCTGAGAGCTGAGAGCTGACCATGTGCT 882
    |||||
```

```
DB 742 GGCACCCCGGCTTGAAGCTCCAGCTGCAAGTCCGAGCTGAGGAGACTGGACCAAGTGT 801
    |||||
QY 883 AGAGATTTGTGTGAGAGACTTGAAGGCCACACC 914
    |||||
DB 802 GGAGATCTGTGTGAGAGCTTGAAGAGAGGCC 833
    |||||
```

## RESULT 3

```
US-09-925-302-153
/ Sequence 153, Application US/09925302
/ Publication No. US20030064072A9
/ GENERAL INFORMATION:
/ APPLICANT: Rosen et al.
/ TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
/ FILE REFERENCE: PA104
/ CURRENT APPLICATION NUMBER: US/09/925,302
/ PRIOR FILING DATE: 2001-08-10
/ PRIOR APPLICATION NUMBER: PCT/US00/05918
/ PRIOR FILING DATE: 2000-03-08
/ PRIOR APPLICATION NUMBER: 60/124,270
/ NUMBER OF SEQ ID NOS: 896
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 153
/ LENGTH: 2038
```

```
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (1490)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: misc feature
/ LOCATION: (1508)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: misc feature
/ LOCATION: (1979)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: misc feature
/ LOCATION: (1992)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: misc feature
/ LOCATION: (2010)
/ OTHER INFORMATION: n equals a,t,g, or c
/ US-09-925-302-153
```

```
Query Match      20.6%; Score 428.4; DB 13; Length 2038;
Best Local Similarity 74.3%; Pred. No. 6.9e-98;
Matches 559; Conservative 0; Mismatches 181; Indels 12; Gaps 1;
```

```
QY 175 CGCCGCCCTGGACAGATGTTTCTAGAGGGCTGAAGAGAAATATGTGACCGAGAGA 234
    |||||
DB 82 CGGGGGTCAGAGACACAATGTTTGCAGAGACTGAAGAGAAATGTTGGCCACGAGGA 141
    |||||
QY 235 AGAGATGAGAGGT-----TTGGCACTGTCCTTCTTATGCTTGCAGGACA 282
    |||||
DB 142 AGACGTGAGGGGGCCCTGGCCGCTTGAAGACAGTCTCTCATACGCTTGCAGGGCA 201
    |||||
QY 283 GTCACTCTGGAGATGTCCTTGTCAAGCTCCAGCTCTGTCAATGCTATGTGAGGCCAA 342
    |||||
DB 202 GTGCTCTGACATGTCCTTGTGTAAGTTGACGCTTTGCAATGCTTGTGAGCCCA 261
    |||||
QY 343 TCTTCCCGCTCGGTCTCATTCGCCAACACAGTCCGGCAGATCCAGAGAGAAATGAGCA 402
    |||||
DB 262 TCTGTGCGGCTCAGTCTCTCATTCGCCAACACAGTCCGGCAGATCCAGAGAGAAATGAGCA 321
    |||||
QY 403 GGAATGTTGTGGATGGAGTGGACATCCCGAATGTAATGCGGCACAGTGAAGCCCT 462
    |||||
DB 322 GGAATGGAAGTGGCGGACAGTGGACATCCCGCAGCTGCGAGACGGGCGCTTGCACGCTT 381
    |||||
QY 463 GGTGTCCACAGATCTGTGTGTACAGTGAAGGGGAGCTGAAGAAAGAGACCCCTGCTCC 522
    |||||
DB 382 GGTGTCCACAGATCTGTGTGTACAGTGAAGGGGAGCTGAAGAAAGAGACCCCTGCTCC 441
    |||||
```



```
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (75) ..(797)
US-10-076-069-3
```

```
Query Match      20.5%; Score 426.8; DB 14; Length 2082;
Best Local Similarity 74.2%; Pred. No. 1.8e-97;
Matches 558; Conservative 0; Mismatches 182; Indels 12; Gaps 1;
```

```
QY      175 CGCCCCCTTGACAGATGTTTGTAGAGGGCTGAAGAGAAATATGTGTACCAAGAGAGA 234
      59 CGGGGTCAAGACACAATGTTTGACAGAGACTGAAGAGAAATGTTGTGCGACAGAGGA 118
      235 AGGAGTAGAGGGT-----TTGGCACTGTCCCTTCTCTATAGCTTGAAGCGGACA 282
      119 AGACGTGAGAGGAGAGCCCTGGCCGGCTTGAAGACAGTCTCTTACATACGCTTGCAGCGGCA 178
QY      283 GTGACTCTTGAGACATGTCCTTGTCAAGCTTCAGCTCTGTACATGCTAGTGAAGCCCAA 342
      179 GTGCTCTCTGGAATGTCCTGTGTAAGTTGCAAGCTTTCACATGCTTGTGAGGCCAA 238
QY      343 TCTCTGCGCTCGGTCTCTATCGGCAACACAGTCCGACATCCAGAGAGAAATGAGCCA 402
      239 CCTGTGCGCTCACTCTCACTTGCACACAGGTCGCGACATCCAGAGAGATGACGCA 298
QY      403 GGATGTGTGTGATGGATGGACCCGAGAAATGTAGATCGGGCACCACTTGAAGCGCT 462
      299 GGATGGAGACGTGGCGCACAGTGGACCCGAGCTGACAGACGGGCGCGCTCGACCGCTT 358
      463 GGTGTCCACAGAGATCTGTGTGTGTACAGTGAAGGGAGCTGAGAGAGAGACCCCTGCTCC 522
      359 GGTCTCAAGAGATCTGTGTGTGTGTGAGGCTGAGGCGGAGAGAGGGGCAATCTGCTCC 418
QY      523 TGAATGGAAGATGCTCCCTTGCAAACTCGGTTTCCGAGCTCCCACTGTGGCTCAGC 582
      419 TGGCTTGGGGAGCGGCGACACAGAGGTCAGATTTTGACCTTGGCCAGTCACTCAGC 478
QY      583 ACCAGGGCAAGAGACCTCTGAGAGCAGCTCTGGGAGATGAGACAGCCCAAGAAAAACAG 642
      479 ACAGGCAACAGAGACCTCTGAGAGCAGCGCTTGGAGATGAGATGGCCCTGAGAAAAACAG 538
QY      643 GGAAGACTTTCAGAAATGCACTGAGACAGATATTTGAGACCTTGAGAGACAAAACTCCAG 702
      539 AGGAAGCTTTCAGAAATGCACTGAGACAGATATTTGAAACCTTGAGAGACAAAACTCCAG 598
QY      703 TTCAGTGAAGAACTCTTCTCAGATGTGACAGCTCTCTACTATGACCTGAGACAGTCT 762
      599 CTGCATGGAAGAGCTGTTCTCAGACGTGACAGCCCTTACTACGACTTGAACAGATACT 658
QY      763 AACAGGAATGATGATGAGGAGCAAGTCACTCTGCAATGAGCTTGAAGGGCTTTGCTGC 822
      659 GAAGAGCATGATGAGGGGGTGCAGGCGGGGCCCTTGGAAGGGCTCGAGGGCTTTGCTGC 718
QY      823 AGCAGCCCTCTCTCCAGTTCCACTTGAAGTCTGACCTGAGCTGAGCTGAGACCATGTGT 882
      719 GGCACACCCAGGCGCTTACGTCCAGCTGACAGTCCGACCTGGGCGAGCTGACACAGTGT 778
QY      883 AGAGATTCTGTGAGAGACTGAGAGGCCACCC 914
      779 GGAGATCTGTGAGAGACTGAGAGAGGCC 810
      Db
      779 GGAGATCTGTGAGAGACTGAGAGAGGCC 810
```

```
RESULT 6
US-10-362-893-5
; Sequence 5, Application US/10362893
; Publication No. US20030211525A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Michael G.
```

```
; APPLICANT: Jung, Kenneth
; TITLE OF INVENTION: GENES EXPRESSED IN THE CELL CYCLE
; FILE REFERENCE: PB-0015 USN
; CURRENT APPLICATION NUMBER: US/10/362,893
; CURRENT FILING DATE: 2003-02-25
```

```
; PRIOR APPLICATION NUMBER: US 60/229,253
; PRIOR FILING DATE: 2000-08-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 2447
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 225657.4
US-10-362-893-5
```

```
Query Match      20.4%; Score 425.2; DB 13; Length 2447;
Best Local Similarity 74.1%; Pred. No. 5.1e-97;
Matches 557; Conservative 0; Mismatches 183; Indels 12; Gaps 1;
```

```
QY      175 CGCCCCCTTGACAGATGTTTGTAGAGGGCTGAAGAGAAATATGTGTACCAAGAGAGA 234
      75 CGGGGTCAAGACACAATGTTTGACAGAGACTGAAGAGAAATGTTGTGCGACAGAGGA 134
QY      235 AGGAGTAGAGGGT-----TTGGCACTGTCCCTTCTCTATAGCTTGAAGGCCCAA 282
      135 AGACGTGAGAGGAGAGCCCTGGCCGGCTTGAAGACAGTCTCTTACATACGCTTGCAGCGGCA 194
QY      283 GTGACTCTTGAGACATGTCCTTGTCAAGCTTCAGCTCTGTACATGCTAGTGAAGCCCAA 342
      195 GTGCTCTCTGGAATGTCCTGTGTAAGTTGCAAGCTTTCACATGCTTGTGAGGCCCAA 254
QY      343 TCTCTGCGCTCGGTCTCTATCGGCAACACAGTCCGCGAGATCCAGAGAGAAATGAGCCA 402
      255 CCTGTGCGCTCACTCTCACTTGCACACAGGTCGCGACATCCAGAGATCCAGAGAGATGACGCA 314
QY      403 GGATGTGTGTGATGGATGGACCCGAGAAATGTAGATCGGGCACCACTTGAAGCGCT 462
      315 GGATGGAGACGTGGCGCACAGTGGACCCGAGCTGACAGACGGGCGCGCTCGACCGCTT 374
QY      463 GGTGTCCACAGAGATCTGTGTGTGTACAGTGAAGGGAGCTGAGAGAGAGACCCCTGCTCC 522
      375 GGTCTCAAGAGATCTGTGTGTGTGTGAGGCTGAGGCGGAGAGAGGGGCAATCTGCTCC 434
      523 TGAATGGAAGATGCTCCCTTGCAAACTCGGTTTCCGAGCTCCCACTGTGGCTCAGC 582
      435 TGGCTTGGGGAGCGGCGACACAGAGGTCAGATTTTGACCTTGGCCAGTCACTCAGC 494
QY      583 ACCAGGGCAAGAGACCTCTGAGAGCAGCTCTGGGAGATGAGACAGCCCAAGAAAAACAG 642
      495 ACAGGCAACAGAGACCTCTGAGAGCAGCGCTTGGGAGATGAGATGGCCCTGAGAAAAACAG 554
QY      643 GGAAGACTTTCAGAAATGCACTGAGACAGATATTTGAGACCTTGAGAGACAAAACTCCAG 702
      555 AGGAAGCTTTCAGAAATGCACTGAGACAGATATTTGAAAGCTTGAGACCTTGAAGAAAACTCCAG 614
QY      703 TTCAGTGAAGAACTCTTCTCAGATGTGACAGCTCTCTACTATGACCTGAGACAGTGT 762
      615 CTGCATGGAAGAGCTGTTCTCAGACGTGACAGCCCTTACTACGACTTGAACAGATACT 674
QY      763 AACAGGAATGATGATGAGGAGCAAGTCACTCTGCAATGAGCTTGAAGGGCTTTGCTGC 822
      675 GAAGAGCATGATGAGGGGGTGCAGGCGGGGCCCTTGGAAGGGCTCGAGGGCTTTGCTGC 734
QY      823 AGCAGCCCTCTCTCCAGTTCCACTTGAAGTCTGACCTGAGCTGAGCTGAGACCATGTGT 882
      735 GGCACACCCAGGCGCTTACGTCCAGCTGACAGTCCGACCTGGGCGAGCTGACACAGTGT 794
QY      883 AGAGATTCTGTGAGAGACTGAGAGGCCACCC 914
      795 GGAGATCTGTGAGAGACTGAGAGAGGCC 826
      Db
```

```
RESULT 7
US-10-037-270-773
; Sequence 773, Application US/10037270
```

```

; Publication No. US20030104529A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Weinman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunding
; APPLICANT: Wang, Dunrui
; APPLICANT: Tillinghast, John
; APPLICANT: Dimanac, Radoje T.
; TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
; FILE REFERENCE: 784CIP28
; CURRENT APPLICATION NUMBER: US/10/037,270
; PRIOR FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pt_fl_genes Version 1.0
; SEQ ID NO 773
; LENGTH: 2629
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (73)..(798)
US-10-037-270-773

Query Match      20.4%; Score 425.2; DB 15; Length 2629;
Best Local Similarity 74.1%; Pred. No 5.4e-97;
Matches 557; Conservative 0; Mismatches 183; Indels 12; Gaps 1;

QY 175 CGCCCGCTTGACAGATGTTTGTCTAGAGGGCTGAAGAGAAATATGTGATCCAGAGAGA 234
DB 57 CGCGGGTCAGAGACACATGTTTGCACGAGACTGAAGAGAAATGTTGGCCACAGGA 116
QY 235 AGAGTAGAGGGT-----TTTGGCACTGTCCCTTCTATAGCTTGACAGGACA 282
DB 117 AGACGTGAGAGAGCCCTGCGGCTTGAAGACAGTGTCTCATACAGCTTGACAGGACA 176
QY 283 GTCACTCTGACATGTCCTTGTCAAGCTCAGCTCTGACATGCTAGTGGAGCCAA 342
DB 177 GTGCTCTGACATGTCCTTGTGAAGTTGACGCTTTCGACATGCTTGTGAGCCAA 236
QY 343 TCTCTCCGCTCGGTCTCTATCCGCAACACAGTCCGACAGATCCAGAGAGAAATGAGCCA 402
DB 237 CCGTGTCCGCTCAGTCTCATTTGCCAACAAGGTCCGACAGATCCAGAGAGATGACGCA 296
QY 403 GGATGTGTGTGATGAGGATGGGATGCCACCCAGAAATGTAAGTCCGGGACCAAGTTGAAGCCT 462
DB 297 GGAATGGAGCTGGCGACAGATGGACCCAGGCTGACAGAGCGGCGCGCTGACCGCTT 356
QY 463 GGTGTCAACAGAGATCTGTGTGTGTAAGTGAAGGGAGCTGAGAGAGACACCTGCTCC 522
DB 357 GGTCTCAACGAGATCTGTGTGTGTAAGTGAAGGGAGCTGAGAGAGACACCTGCTCC 416
QY 523 TGAATCGGAAGATGCTCTCTGCAAAATCGGTTTCCGAGCTCCCATGTTGGCTCAGC 582
DB 417 TGGCTTGGGGAGAGGGCACACAGAGTCCAGTTTCTGACTTTGGCCAGTCACTCAGC 476
QY 583 ACCAGGGCAAGAGACCTCAAGACAGCTCTGGGAGATGAGCAGCCCAAGAGAAACAG 642

```

```

DB 477 ACAGGACCAAGGACCTGACAGAGCGCCCTGGAGATGATGCGCCCTCGAGAAAACAG 536
QY 643 GGGAGCTTTACAGAGTCACTGACCAAGATATTTGAGACCTTGAGAAACAAAATCCAG 702
DB 537 AGGAGCTTTACAGAGTCACTGATCATATTTGAACGCTGAGAGACTAAACCCAG 596
QY 703 TTCAATGAGAGAACTTTTCAAGATGTGACAGCTCTCTATCTATGACCTGACACAGTGT 762
DB 597 CTGATGAGAGAGCTGTCTTCAAGAGCTGACAGCCCTTACTGACCTGACACAGTACT 656
QY 763 AACAGAAATGATGATGAGGAGCAAGTCCAGTCTGCAATGAGCTTGAAGGCTTGCTGC 822
DB 657 GACAGGATGATGAGGAGGAGTCCAGGCGGCGCCCTGCAAGGCTTGAGGCTTGCTGC 716
QY 823 AGCCACCTCTCTCCAGTTCCAGTTGCAAGTGTGACCTGCTGAGCTGACATGTGT 882
DB 717 GGCACCCCAAGGCTTCACTGCTCAGCTGCAAGTCCAGCTGGGCGAGCTGACCAAGTGT 776
QY 883 AGGATTTCTGAGAGAGCTGAGAGCCACC 914
DB 777 GAGATCTGTGAGAGCTGAGAGAGCC 808

```

## RESULT 8

US-10-117-722-773

; Sequence 773, Application US/10117722

; Publication No. US20030219744A1

; GENERAL INFORMATION:

; APPLICANT: Tang, Y. Tom

; APPLICANT: Liu, Chenghua

; APPLICANT: Asundi, Vinod

; APPLICANT: Zhang, Jie

; APPLICANT: Dimanac, Radoje T.

; TITLE OF INVENTION: No. US20030219744A1el Nucleic Acids and

; FILE REFERENCE: 784CIP28CIP

; CURRENT APPLICATION NUMBER: US/10/117,722

; PRIOR FILING DATE: 2002-04-04

; PRIOR APPLICATION NUMBER: 09/620,312

; PRIOR FILING DATE: 2000-07-19

; PRIOR APPLICATION NUMBER: 09/552,317

; PRIOR FILING DATE: 2000-04-25

; PRIOR APPLICATION NUMBER: 09/488,725

; NUMBER OF SEQ ID NOS: 1104

; SOFTWARE: pt\_fl\_genes Version 1.0

; SEQ ID NO 773

; LENGTH: 2629

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (73)..(798)

US-10-117-722-773

Query Match 20.4%; Score 425.2; DB 16; Length 2629;

Best Local Similarity 74.1%; Pred. No 5.4e-97;

Matches 557; Conservative 0; Mismatches 183; Indels 12; Gaps 1;

```

QY 175 CGCCCGCTTGACAGATGTTTGTCTAGAGGGCTGAAGAGAAATATGTGATCCAGAGAGA 234
DB 57 CGCGGGTCAGAGACACATGTTTGCACGAGACTGAAGAGAAATGTTGGCCACAGGA 116
QY 235 AGAGTAGAGGGT-----TTTGGCACTGTCCCTTCTATAGCTTGACAGGACA 282
DB 117 AGACGTGAGAGAGCCCTGCGGCTTGAAGACAGTGTCTCATACAGCTTGACAGGACA 176
QY 283 GTCACTCTGACATGTCCTTGTCAAGCTCAGCTCTGACATGCTAGTGGAGCCAA 342
DB 177 GTGCTCTGACATGTCCTTGTGAAGTTGACGCTTTCGACATGCTTGTGAGCCAA 236
QY 343 TCTCTCCGCTCGGTCTCTATCCGCAACACAGTCCGACAGATCCAGAGAGAAATGAGCCA 402

```



```

1  APPLICANT: Chenault, Ruth A.
2  APPLICANT: Xu, Jitengchun
3  APPLICANT: Fanger, Gary R.
4  APPLICANT: Harlocker, Susan L.
5  APPLICANT: McNeill, Patricia D.
6  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
7  TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
8  FILE REFERENCE: 210121.486C7
9  CURRENT APPLICATION NUMBER: US/10/212,677
10 CURRENT FILING DATE: 2002-08-02
11 NUMBER OF SEQ ID NOS: 288
12 SOFTWARE: FastSeq for Windows Version 4.0
13 SEQ ID NO 167
14 LENGTH: 396
15 TYPE: DNA
16 ORGANISM: Homo sapiens
17 US-10-212-677-167

```

Query Match	9.1%;	Score 190;	DB 15;	Length 396;
Best Local Similarity	74.7%;	Pred. No. 1.1e-37;		
Matches 257;	Conservative 0;	Mismatches 75;	Indels 12;	Gaps 1;

0y CGCCCGCCTGGA CAGAGTGTTC TCTAGAAGGCTGA AGAGAAATGTGTG ATCCAGGAGA 234

175 CGCCCGCCTGGA CAGAGTGTTC TCTAGAAGGCTGA AGAGAAATGTGTG ATCCAGGAGA 234

52 CGCCCGCCTGGA CAGAGTGTTC TCTAGAAGGCTGA AGAGAAATGTGTG ATCCAGGAGA 111

Db

235 AGGAGTAGAGGGT-----TTTGCACTCTCCCTTCCTATAGCCTGCAGCGACA 282  
 112 AGAGCTGCAGCGAGAGCCCTGGCCGGCTTGAAAGACAAGTCTCTCATACAGCTTGCAGCGGCA 171

0y 28 GTGACCTCCGAGACATGTCCCTTGTCAAGCTCAGCTCTGTACACATGCTAGTGAGCCCA 342  
Db 172 GTCCCTCCGAGACATGTCTCTGTGGAAGTTGCAGCTTTGGCCACATGCTTGTGAGCCCA 231

343 TCTCTGCCCCCTCGGTCCTCATCGCCACACAGATCCGCGAGGAGGAAATAGCCA 402  
 07 TCTCTGCCCCCTCGGTCCTCATCGCCACACAGATCCGCGAGGAGGAAATAGCCA 402  
 Db 232 TCTGTGCCGCTCAGTCTCTCATTTGCCAACACGGTCCGGCAGATCCAAAGAGGATGACGA 291

Db 292 GGATGGGACGTGGCGCACAGTGGCCACCCCGAGCTGCAGAGCGGGCGCCGCTCGACCGCTT 351

47 GGTCTCCACGAGATCCTGTGCCCGTCACAGGTGGGGCAAGAGG 395

RESULT 12  
US-10-361-811-167  
: Sequence 167, Application US/10361811

```

: GENERAL INFORMATION:
:
: APPLICANT: Fanger, Gary R.
:
: APPLICANT: Fling, Steven P.

```

```

; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.484C8
; CURRENT APPLICATION NUMBER: US/10/361,811

```

```

;      NUMBER OF SEQ ID NOS: 293
;      SOFTWARE: FastSeq for Windows Version 4.0
;      SEQ ID NO 167

```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-361-811-167

```

Query Match	9.1%;	Score 190;	DB 16;	length 356;
Best Local Similarity	74.7%;	Pred. No. 1.1e-37;		
Matches 257;	Conservative	0;	Mismatches 75;	Indels 12;
				Gaps 1;

175 CGGCCCGCCCTGGACAGAGATGTTTGTCTAGAGAGGCTGTAAGAGAGAAATATGGTGACCCAGGAGGA 234

Db 52 CGGGGTCAGGACACAAATGTTTGACGAGGACTGAAGAGAAATGTGTGGCCACGAG 111

Qy 235 AGGAGTAAAGGGT-----TTTGGCAGCTGTCCCTTCCATATAGCTTCGACGGCA 282

Db 112 AGACGGGAGGAGAGGCCCTGGCCGGCTTAAAGACAGTGTCTTCATACAGCTTCGAGGGCA 171

Qy 283 GTACACTCTCGACATGTCCCTTGTCAAGCTCAGCTCTGTGCATCTGTAAGGCCAA 342

Db 172 GTCGCTCTTGGAACATGTCTGTGTGAATTGACAGTTTGGCAATGCTTGTGGAGCCAA 211

Qy 343 TCTCTCCGCGCTGGGTCTTCATGTGCCAACAAGTCCGCGAATCCAGAGGAAATGACCA 402

Db 232 TCTGTCCGCTCAGTCTTCATTTGCCAACGCGTCCGAGATCCAAAGGAGATGACGA 291

Qy 403 GGATGTGTGTGGCATGTGGATGGCACCACAAATGTAGATTCGGGACCAAGTTAAAGCCT 462

Db 292 GGATGGGAAGTGGGCGACAGTGGCACCACCAAGCTGTGAGAGGGCGCGCTTCGACCGCTT 351

Qy 463 GGTGTCCACAGAGATCTCTGTGTCTGACAGTAAAGGAGACGTGAG 506

Db 352 GGTCTCCACGAGATCTGTGTGCGGACGCTGTGGGCGCAAGAG 395

RESULT 13  
US-10-369-186-167

Publication No. US20030232056A1  
GENERAL INFORMATION:  
APPLICANT: Fanger, Gary R.  
ATTORNEY: Fanger, Gary R.

```

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; OF ENDOMETRIAL HYPERPLASIA AND PREVENTION OF ENDOMETRIAL
; CANCER
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.484C9
; CURRENT ABSTRACT NUMBER: 115/110/268 186

```

```

; CURRENT FILING DATE: 2003-02-14
;
; NUMBER OF SEQ ID NOS: 293
;
; SOFTWARE: FastSeq for Windows Version 4.0
;
; SEQ ID NO: 167

```

```

;
; LENGTH: 396
;
; TYPE: DNA
;
; ORGANISM: Homo sapiens
HS-10-369-186-167

```

Query Match	9.1%	Score 190;	DB 16;	Length 396;
Best Local Similarity	74.7%	Pred. No. 1.1e-37;		
Matches 257; Conservative	0;	Mismatches 75;	Indels 12;	Gaps 1

**OY**

175 CGCCCGCTGGACAGATGTTTGCTAGACGCCTGAAGAATAAATTGTGCACAAGAGA   234  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
**D6**

52 CCGCGGTCAGGACA CAATGTTTTGCCAGAGACTGAAGAGAAATGTGTGCCCAAGAGA   111

Oy 235 AGGAGTAGAGGGT-----TTTGACACTGTCCCTTCCTATAGCCTGCAGCGACA 282  
||| |||||  
Db 112 AGACGTGAGGGGACCCTGGCCCGGCTTGAAGAACAAGTGTCTCATACAGCCTGCAGCGGCA 171

Oy	283 GTCACCTCCTGAGACAATGTCCCTTGTCAAGCTCCAGCTTGTCACATCGTAAGTAGGCCAA	342
Dd	172 GTGCCTCCGGACAATGTTCTCTGTGGGAATTGCAGCTTTGCCACACATGCTTTGTGGAGCCCCA	231

OY      343 TCTCTGCGCGTGGTCCTCATCGCCAAACAGTCCGGCAGATCCAAGAGGAATTAGGCCA 402  
| | | | |  
DB    232 TCCTGTGCGGCTCACTCCTCATTTGCACAACGTCGGCAGATCCAAGAGGAGATGACGCA 291

Dy 403 GGATGGTGTGTGCGCATGGATGGACCCAGAACTAGATCGGCACCAATTGAACGCT 462  
| | | | | | | | | | | | | | | | | | | | |  
Db 292 GGATGGACGTGGCGCACAGTGACCACCGAGCTGCAGACGGGCCGCTCGACCGCTT 351

**Qy** 463 GGTGTCACAGAGATCCTGTGTCGTAAGTGAGGGAGCTGAGG 506  
|||||  
**Db** 352 GGTTCCACGAGATCCTGTGCGCTGCACGCTGGGGGCAAGAGG 395  
|||||

## RESULT 14



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 10, 2004, 10:57:36 ; Search time 5666 Seconds  
(without alignments)  
10973.007 Million cell updates/sec

Title: US-10-076-069-1

Perfect score: 2082

Sequence: 1 cccgcgcgcgcgcctccacc.....atgaaaaaaaaaaaaaaaaa 2082

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

EST.\*  
1: em\_estba:\*  
2: em\_estbum:\*  
3: em\_estlin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_man:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vtl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1883.8	90.5	1985	11	AK010535 Mus muscu
2	1883.8	90.5	1996	11	AK032980 Mus muscu
3	890.2	42.8	1180	13	BU511608 AGENCOURT
4	849.2	40.8	989	9	AV000968 AV000968

5	810.4	38.9	836	13	BU054465	BU054465 UI-M-FD0-
6	809.8	38.9	937	14	CB183964	CB183964 AGENCOURT
7	798	38.3	927	13	BO950078	BO950078 AGENCOURT
8	780.6	38.0	920	13	BU148336	BU148336 AGENCOURT
9	785.8	37.7	926	13	BU151870	BU151870 AGENCOURT
10	755.4	36.3	994	10	BB609465	BB609465 BB609465
11	748.4	35.9	967	13	BO936878	BO936878 AGENCOURT
12	746	35.8	922	13	BO713776	BO713776 AGENCOURT
13	741	35.6	830	12	BI661691	BI661691 603305632
14	721.4	34.6	800	12	BI153019	BI153019 602918211
15	714	34.3	714	29	AY403475	AY403475 Mus muscu
16	697.8	33.5	809	13	BO572821	BO572821 UI-M-FD0-
17	694.8	33.4	826	13	BU076179	BU076179 UI-M-FR0-
18	687.8	33.0	815	14	CA752039	CA752039 UI-M-FD0-
19	681.8	32.7	824	14	CA752149	CA752149 UI-M-FD0-
20	677.8	32.6	748	12	BI689943	BI689943 603310777
21	677.8	32.6	876	12	BI685217	BI685217 603310227
22	677.8	32.5	753	12	BI685932	BI685932 603309232
23	676.4	32.5	936	13	BU512943	BU512943 AGENCOURT
24	672.8	32.3	911	13	BO886760	BO886760 AGENCOURT
25	669.4	32.2	757	12	BI653556	BI653556 603300591
26	668	32.1	679	10	BG69072	BG69072 H3072D09-
27	665.6	32.0	713	12	BI665883	BI665883 603286912
28	664.2	31.9	682	10	BB055758	BB055758 BB055758
29	661.6	31.8	914	14	CB587874	CB587874 AGENCOURT
30	645.8	31.0	649	14	CB057494	CB057494 NISC J805
31	636.6	30.6	724	14	CA751730	CA751730 UI-M-FD0-
32	620.4	29.8	898	13	BU515897	BU515897 AGENCOURT
33	594.6	28.6	724	10	BF607870	BF607870 MY1.00081
34	588.8	28.3	679	13	BY726193	BY726193 BY726193
35	585.6	28.1	804	10	BE573436	BE573436 601332611
36	573.8	27.6	648	14	CF173487	CF173487 BO924D06-
37	573	27.5	730	12	BI294797	BI294797 UI-R-DK0-
38	571.8	27.5	646	14	CF916191	CF916191 BO991D06-
39	568.2	27.3	726	12	BI077734	BI077734 602871683
40	564.4	27.1	589	10	AM547757	AM547757 L0026D04-
41	552	26.5	552	12	BI663600	BI663600 603288096
42	548.2	26.3	906	10	BF144504	BF144504 601790146
43	544.4	26.1	600	14	CA528387	CA528387 8072.96 M
44	528.6	25.4	568	12	BI666485	BI666485 603288772
45	525	25.2	613	29	CG586964	CG586964 OST236075

## ALIGNMENTS

RESULT 1  
AK010535  
LOCUS  
DEFINITION Mus musculus ES cells cDNA, RIKEN full-length enriched library,  
clone:2410018C03 product:HEMATOPOIETIC PROGENITOR PROTEIN, full  
insert sequence.  
ACCESSION AK010535  
VERSION AK010535.1 GI:12846052  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
REFERENCE  
1 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PUBMED 10349636  
REFERENCE  
2 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,  
TITLE Itch,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
JOURNAL Normalization and subtraction of cap-trapper-selected cDNAs to  
MEDLINE 20493374  
PUBMED 11042159



Db 886 TTGTGACCCAGACAGATAGCACTTGTCTTAAGAGGGGCTGTGGTCTTGAAGCTATT 945  
QY 1022 ATCTTTTGTGTGACATTTGACATCTGTGAGAGATGGTGTGCACAGCTATGTCTATGTC 1081  
Db 946 ATCTTTTGTGTGACATTTGACATCTGTGAGAGATGGTGTGCACAGCTATGTCTATGTC 1005  
QY 1082 TATTTTCAATTAAGTGGTGAACCTTTCTAAATTTAAGTTTATATGTTTGGGCAATAT 1141  
Db 1006 TATTTTCAATTAAGTGGTGAACCTTTCTAAATTTAAGTTTATATGTTTGGGCAATAT 1065  
QY 1142 TTTGTCTTAAGATATATTTTAACTTTTATATCTTTAGATTTTTCAGCTATTTTCT 1201  
Db 1066 TTTGTCTTAAGATATATTTTAACTTTTATATCTTTAGATTTTTCAGCTATTTTCT 1125  
QY 1202 TAAAGATATATTTTCTACAAACATCTCTGCTGCTACATTTGAAACATTTATACCTA 1261  
Db 1126 TAAAGATATATTTTCTACAAACATCTCTGCTGCTACATTTGAAACATTTATACCTA 1185  
QY 1262 AATACGATTTGGTGTGATTTTAAAGTTTAAATAGAAACCTTTTGTAGTGTCT 1321  
Db 1186 AATACGATTTGGTGTGATTTTAAAGTTTAAATAGAAACCTTTTGTAGTGTCT 1245  
QY 1322 CTACACTCCCAAGCACTGTAAATGTAGCCGCGCGGGTGTTTACATGAGAGCTCCAGT 1381  
Db 1246 CTACACTCCCAAGCACTGTAAATGTAGCCGCGCGGGTGTTTACATGAGAGCTCCAGT 1305  
QY 1382 ATGGTCTACATTTAGTAGAGCTTTGAAAGAACCATGACAGCTCCACCTCCCTCAGT 1441  
Db 1306 ATGGTCTACATTTAGTAGAGCTTTGAAAGAACCATGACAGCTCCACCTCCCTCAGT 1365  
QY 1442 GGGTCTGCTGTGGCGGATGGAGACTCTTCTTACAGCCCTGTGCGAGATGGCTTATTT 1501  
Db 1366 GGGTCTGCTGTGGCGGATGGAGACTCTTCTTACAGCCCTGTGCGAGATGGCTTATTT 1425  
QY 1502 ATGCTATTTATATGTAAATGSCACATGAAAGCTAAGGCTTACCTCGGAATCCCAACA 1561  
Db 1426 ATGCTATTTATATGTAAATGSCACATGAAAGCTAAGGCTTACCTCGGAATCCCAACA 1485  
QY 1562 CCAAGTTCTTCAAGGAGCTGCTGTGAGAGGAGCTTATGACAGGCTTGTCTTGGCCATCA 1621  
Db 1486 CCAAGTTCTTCAAGGAGCTGCTGTGAGAGGAGCTTATGACAGGCTTGTCTTGGCCATCA 1545  
QY 1622 CTGTCTGTGTTCCAGCCAGCCACATGTGACATGAGAGCATGACATGCGCCAGACCCAG 1681  
Db 1546 CTGTCTGTGTTCCAGCCAGCCACATGTGACATGAGAGCATGACATGCGCCAGACCCAG 1605  
QY 1682 CACCAATGCTGCATGTCAAGTGTGACGTGAGAGACCACTGGCTCCAGGCTGTGCTCA 1741  
Db 1606 CACCAATGCTGCATGTCAAGTGTGACGTGAGAGACCACTGGCTCCAGGCTGTGCTCA 1665  
QY 1742 GAGAGGGTGTGAGTCTTACGTGTGCTGTGGGGGGGAGCAGCGGTGAACCTGTCTTGTCT 1801  
Db 1666 GAGAGGGTGTGAGTCTTACGTGTGCTGTGGGGGGGAGCAGCGGTGAACCTGTCTTGTCT 1725  
QY 1802 TTTTAAATGTGTCTTGGACGTTTAAAGTTTAAACAAATCCGACTCCATATGATTTAGGG 1861  
Db 1726 TTTTAAATGTGTCTTGGACGTTTAAAGTTTAAACAAATCCGACTCCATATGATTTAGGG 1785  
QY 1862 CTCTCTCAACCTGGGGGGGCCCCCTATGCTGTGCTTGAATTCGAAGCTTTGGTACTCG 1921  
Db 1786 CTCTCTCAACCTGGGGGGGCCCCCTATGCTGTGCTTGAATTCGAAGCTTTGGTACTCG 1845  
QY 1922 GCACTGTGACAGCTCCACCCCATGATCTTTTGTGTCTTGTGCTTTTGGACTTC 1981  
Db 1846 GCACTGTGACAGCTCCACCCCATGATCTTTTGTGTCTTGTGCTTTTGGACTTC 1905  
QY 1982 CCAACTGAGCCTTAAGTTTATTTATATGTGCTTCAATATCAACAATGTAAACCTCAC 2041  
Db 1906 CCAACTGAGCCTTAAGTTTATTTATATGTGCTTCAATATCAACAATGTAAACCTCAC 1965  
QY 2042 TTTATTTAAAGTATCCAGC 2060  
|||||

Db 1966 TTTATTTAAAGTATCCAGC 1984  
RESULT 2  
AK032980  
LOCUS  
DEFINITION  
AK032980 1966 bp mRNA linear HTC 18-SEP-2003  
Mus musculus 12 days embryo male wolffian duct includes surrounding  
region cDNA, RIKEN full-length enriched library, clone:6720484A10  
product:HEMATOPOIETIC PROGENITOR PROTEIN, full insert sequence.  
ACCESSION  
AK032980.1 GI:26328740  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
AUTHORS  
1 Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
99279253  
MEDLINE  
PUBMED  
10349636  
REFERENCE  
AUTHORS  
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtrahction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
20499374  
MEDLINE  
PUBMED  
11042159  
REFERENCE  
AUTHORS  
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P.,  
Komno, H., Akiyama, J., Nishi, K., Kikunishi, T., Teshiro, H., Itoh, M.,  
Sumi, N., Ieshi, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
AUTHORS  
4 The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)  
5 The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 1996)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T.,  
Hori, F., Imotani, K., Ishii, Y., Ito, M., Kagawa, I., Kasukawa, T.,  
Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Komno, H., Koda, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohnato, N.,  
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shitaki, T.,  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Kahira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muramatsu, M. and Hayashizaki, Y.  
Direct Submission  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail:genome-res@gs.c.riken.go.jp),

## COMMENT

URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site for further details.  
URL: <http://genome.gsc.riken.go.jp/>,  
Location/Qualifiers

## FEATURES

## source

1. 1996

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="RANTOM:DB:6720484A10"

/db\_xref="MGI:2396648"

/db\_xref="taxon:10090"

/clone="6720484A10"

/sex="male"

/issue\_type="wolfian duct includes surrounding region"

/clone\_lib="RIKEN full-length enriched mouse cDNA library"

/dev\_stage="12 days embryo"

126. 839

/note="unnamed protein product; HEMATOPOIETIC PROGENITOR PROTEIN (SPTR|O99MP6, evidence: FASTV, 100%ID, 100%length, match=711)

putative"

/codon\_start=1

/protein\_id="BAC28109.1"

/db\_xref="GI:26328741"

/translation="MFARGLKRYGDDEEVEFGTVPSYLSORSLDMVLKLOLC HMLVEPLCRSVLIANTROIEMSDGWHGMAPONVRAVERLVSLEICRYR GAEHEPAPLEADAPLONSYSELPIVSGAGORPOSIMEMSPENRSPKSIDQ IPEETLKNSSVBELEFSDVSSYIDLTVLTMGSGTSSLCNGLEGFAPAAPPS TCKSDLAELDHVEILVET"

1980. 1985

polya\_signal

/note="putative"

1996

polya\_site

/note="putative"

## ORIGIN

Query Match 90.5%; Score 1883.8; DB 11; Length 1996;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1878; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 182 CTGACACGATGTTGCTAGAGGGCTGAAGAGAAATATGTACCAAGAGAGAGTA 241  
DB 117 CCGACACGATGTTGCTAGAGGGCTGAAGAGAAATATGTACCAAGAGAGAGTA 176  
QY 242 GAGGGTTTGGCACTGCTCCTCTATAGCCTGACAGACAGTCACTCCTGACATGTCC 301  
DB 177 GAGGGTTTGGCACTGCTCCTCTATAGCCTGACAGACAGTCACTCCTGACATGTCC 236  
QY 302 CTTGTCAAGTCCAGCTCTGTCAATGTAAGTGAAGCCAAATCTGCGCTGGTCTTC 361  
DB 237 CTTGTCAAGTCCAGCTCTGTCAATGTAAGTGAAGCCAAATCTGCGCTGGTCTTC 296  
QY 362 ATGCCAACACAGTCCGGGACATCCAGAGAGAAATAGCCAGAGATGTGTGGCATGGG 421  
DB 297 ATGCCAACACAGTCCGGGACATCCAGAGAGAAATAGCCAGAGATGTGTGGCATGGG 356  
QY 422 ATGCCAACCCAGATATGATTCGGGACCAAGTTGAAGCCTGGTGTCCACAGAGATCTCG 481  
DB 357 ATGCCAACCCAGATATGATTCGGGACCAAGTTGAAGCCTGGTGTCCACAGAGATCTCG 416  
QY 482 TGTCTTACAGTGGGGAGCTGAGAGAAAGACCTGTCTCTGAACCTGGAAGATGCTCC 541  
DB 417 TGTCTTACAGTGGGGAGCTGAGAGAAAGACCTGTCTCTGAACCTGGAAGATGCTCC 476  
QY 542 TTGCAAAACTCGGTTTCCGAGCTCCCATCGTTGGTCTAGACACCAAGGGGAAAGAACCT 601  
DB 477 TTGCAAAACTCGGTTTCCGAGCTCCCATCGTTGGTCTAGACACCAAGGGGAAAGAACCT 536

QY 602 CAGAGCAGCTCTGGGAGATGACACGCCCAAGAAAAAGGGGAAAGCTTTGAGAAATCA 661  
DB 537 CAGAGCAGCTCTGGGAGATGACACGCCCAAGAAAAAGGGGAAAGCTTTGAGAAATCA 596  
QY 662 CTGACACGATATTTAGACCCCTGGAGAACAAAACTCCAGTTCAGTGAAGAACTCTTC 721  
DB 597 CTGACACGATATTTAGACCCCTGGAGAACAAAACTCCAGTTCAGTGAAGAACTCTTC 656  
QY 722 TCAGATGTGACAGACTCTACTATATGACCTGACACAGTCTAAACAGAAATGATAGTGG 781  
DB 657 TCAGATGTGACAGACTCTACTATATGACCTGACACAGTCTAAACAGAAATGATAGTGG 716  
QY 782 ACCAATCCAGTCTCTGCAATGAGCCTTGAGGCTTTGCTGACGCCACCCCTCTCCAGT 841  
DB 717 ACCAATCCAGTCTCTGCAATGAGCCTTGAGGCTTTGCTGACGCCACCCCTCTCCAGT 776  
QY 842 TCCACTTGCAGAGTCTGACCTGGGTGAGTGCATGTGTGAGATTCTGGTGAAGACC 901  
DB 777 TCCACTTGCAGAGTCTGACCTGGGTGAGTGCATGTGTGAGATTCTGGTGAAGACC 836  
QY 902 TGAGAGGCCACCCCAAGTGGGCTAAGGGTGAAGCCACAGTCCCAATGAGCTCAAGTGG 961  
DB 837 TGAGAGGCCACCCCAAGTGGGCTAAGGGTGAAGCCACAGTCCCAATGAGCTCAAGTGG 896  
QY 962 TGTGTACCCAGAGACAGATTAAGCACTTGCTCTAAGAGGGGCTCTGGCTTTAGCTCAT 1021  
DB 897 TGTGTACCCAGAGACAGATTAAGCACTTGCTCTAAGAGGGGCTCTGGCTTTAGCTCAT 956  
QY 1022 ATCCCTTTGTGATGATGGAATCACTGTGAGAGATGGTGTGATCAAGTATGTCTAGTC 1081  
DB 957 ATCCCTTTGTGATGATGGAATCACTGTGAGAGATGGTGTGATCAAGTATGTCTAGTC 1016  
QY 1082 TATTTTCAATTAGATAGTGAACCTTCTAATAATTAAGTTTATATGTTTGGGCAATAT 1141  
DB 1017 TATTTTCAATTAGATAGTGAACCTTCTAATAATTAAGTTTATATGTTTGGGCAATAT 1076  
QY 1142 TTTGTCTTAAGATATATTTTAACTTTAACTTTAATTTAGATTTTTCAGTATTTTCT 1201  
DB 1077 TTTGTCTTAAGATATATTTTAACTTTTAACTTTTAACTTTTAACTTTTTCAGTATTTTCT 1136  
QY 1202 TAAAGATATATTTTCTACAAACATCCTCTGCTGCTACATTAAGAAATTTAATACCTA 1261  
DB 1137 TAAAGATATATTTTCTACAAACATCCTCTGCTGCTACATTAAGAAATTTAATACCTA 1196  
QY 1262 AATACGATTTGGTGTCAATTTTAAAGTTTAAATGAAGAACTTTTGTGATGAGTCT 1321  
DB 1197 AATACGATTTGGTGTCAATTTTAAAGTTTAAATGAAGAACTTTTGTGATGAGTCT 1256  
QY 1322 CTACACTCCCAAGGCACTGTAAATGTAGCCGGCCGGGTGTTTACATAGAGGCTCCAGT 1381  
DB 1257 CTACACTCCCAAGGCACTGTAAATGTAGCCGGCCGGGTGTTTACATAGAGGCTCCAGT 1316  
QY 1382 ATGATCTACATTTAGTGAAGCTTGAAAAAGAACATGACAGCTCACTGCCCCCTCACT 1441  
DB 1317 ATGATCTACATTTAGTGAAGCTTGAAAAAGAACATGACAGCTCACTGCCCCCTCACT 1376  
QY 1442 GGGTGTCTCTGGGGAGATCGAGAGCTCTTTCTAGCCCGGTGTGAGAGATGGCTTATTT 1501  
DB 1377 GGGTGTCTCTGGGGAGATCGAGAGCTCTTTCTAGCCCGGTGTGAGAGATGGCTTATTT 1436  
QY 1502 ATGCTATTTTATGTATAATGCACTGAAGCTTAAGCTTAACCTCGGAAATCCCAACA 1561  
DB 1437 ATGCTATTTTATGTATAATGCACTGAAGCTTAAGCTTAACCTCGGAAATCCCAACA 1496  
QY 1562 CCAAGTTCTTGAAGGACCTGCTGAAGGAGTGCCTTATGACAGTCTTGTCTTGGCCATCA 1621  
DB 1497 CCAAGTTCTTGAAGGAGTGCCTGTAAGGAGTGCCTTATGACAGTCTTGTCTTGGCCATCA 1556  
QY 1622 CTGCTGTGTTTCCAGCCAGACATGTGACATGAGACATGACATGCCCCAACCACCCAG 1681  
DB 1557 CTGCTGTGTTTCCAGCCAGACATGTGACATGAGACATGACATGCCCCAACCACCCAG 1616

QY 1682 CACCACATGCTCCATGTCAGATGTGTACGTGAGACCACTGGCTCCAGGCTGTGTCTCA 1741  
DB 1617 CACCACATGCTCCATGTCAGATGTGTACGTGAGACCACTGGCTCCAGGCTGTGTCTCA 1676  
QY 1742 GAGAGGGTGTGACATCTCTACGTGTGTCTGGGGGAGACACGGTGAACCTGTGTGTCT 1801  
DB 1677 GAGAGGGTGTGACATCTCTACGTGTGTCTGGGGGAGACACGGTGAACCTGTGTGTCT 1736  
QY 1802 TTTTAAATGTGCTGTGAACGTTTAAAGTTAAACAAATCCGACTCCATATGATTTAGG 1861  
DB 1737 TTTTAAATGTGCTGTGAACGTTTAAAGTTAAACAAATCCGACTCCATATGATTTAGG 1796  
QY 1862 CTCTCCACCTGGGGGTGGCCCTATGCTGTCTGTGATCTCAAGCTTGGTACTCG 1921  
DB 1797 CTCTCCACCTGGGGGTGGCCCTATGCTGTCTGTGATCTCAAGCTTGGTACTCG 1856  
QY 1922 GCACTGTGACATCTCAACCCCAATGATCTTTTGTGTCTGTGTCTTTTGTGAGCTTC 1981  
DB 1857 GCACTGTGACATCTCAACCCCAATGATCTTTTGTGTCTGTGTCTTTTGTGAGCTTC 1916  
QY 1982 CCAACCTGAGCTTAAAGTTTATTTATGCTGCTTCAATATCAATGTAACCTTCAC 2041  
DB 1917 CCAACCTGAGCTTAAAGTTTATTTATGCTGCTTCAATATCAATGTAACCTTCAC 1976  
QY 2042 TTTATTTAAAGTATCCAGC 2060  
DB 1977 TTTATTTAAAGTATCCAGC 1995

RESULT 3  
LOCUS BU511608 1180 bp mRNA linear EST 12-SEP-2002  
DEFINITION AGENCOURT 10129003 NIH\_MGC\_134 Mus musculus cDNA clone  
BU511608  
IMAGE:6506412 5', mRNA sequence.  
BU511608 1 GI:22817841  
EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 1180)  
REFERENCE NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: csapbs-rc@mail.nih.gov  
Tissue Procurement: Dr. David Rowe  
cDNA Library Preparation: Invitrogen Corp.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC Clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.jnl.gov  
Plate: LHAM14068 row: n column: 13  
High quality sequence start: 12  
High quality sequence stop: 719.  
Location/Qualifiers  
1. 1180

FEATURES  
source

1. 1180  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6506412"  
/tissue\_type="undifferentiated limb"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_id="NIH\_MGC\_134"  
/note="Vector: DCMV-SPORE.1; Site 1: EcoRV; Site 2: NotI;  
Cloned unidirectionally. Primer: Oligo dT. Average insert  
size 1.7 Kb. Constructed by Resgen, Invitrogen Corp. Note:  
this is a NIH\_MGC Library."

ORIGIN  
Query Match

42.8%; Score 890.2; DB 13; Length 1180;

Best Local Similarity 89.4%; Pred. No. 3,5e-167;  
Matches 1004; Conservative 0; Mismatches 107; Indels 12; Gaps 4;  
QY 98 GCCCATGCGGACCCCTTTGCTGGGCAAGTACCGATTCGGGGTCCCGGAGAGCTGGCG 157  
DB 42 GCGCATGCGGACCCCTTTGCTGGG---CGTGAACGATCCGGGTGCCGGAAGAG--TGGC 96  
QY 158 TGGGTCTGCTTGCAGCCCGCCGCTGGAAGAGATGTTGCTGAGAGGCTGGAAGAGAA 217  
DB 97 GGGGTCTGCTTGGCGC-----GCGGACAGATGTTTCTGAGAGGCTGGAAGAGAA 150  
QY 218 TATGTGACCAAGAGAAGAGATGAGAGGTTTGGACATGTCCCTTCTTAAGCTTGAC 277  
DB 151 TATGTGACCAAGAGAAGAGATGAGAGGTTTGGACATGTCCCTTCTTAAGCTTGAC 210  
QY 278 CGACATGCTCTCTGACATGTCTCTTGTAAAGTCAAGTCTGTGACATGCTAGTGAAG 337  
DB 211 CGACATGCTCTCTGACATGTCTCTTGTAAAGTCAAGTCTGTGACATGCTAGTGAAG 270  
QY 338 CCCAATCTGCGCGCTCGGTCTCATGCGCAACAGTCCGGGAGATCCAGAGAGAAATG 397  
DB 271 CCCAATCTGCGCGCTCGGTCTCATGCGCAACAGTCCGGGAGATCCAGAGAGAAATG 330  
QY 398 AGCCAGATGCTGTGTGATGAGATGACACCCAGATGATGATGCGGACCAAGTTGAA 457  
DB 331 AGCCAGATGCTGTGTGATGAGATGACACCCAGATGATGATGCGGACCAAGTTGAA 390  
QY 458 CGCCTGTGTCTCAAGATGCTGTGTCTGTACGTAGAGGAGAGTGAAGAGACCTT 517  
DB 391 CGCCTGTGTCTCAAGATGCTGTGTCTGTACGTAGAGGAGAGTGAAGAGACCTT 450  
QY 518 GCTCTGAACTGGAAGTGTCTCTTGTGCAAAATCGGTTTCGAGCTCCCATCGTTGGC 577  
DB 451 GCTCTGAACTGGAAGTGTCTCTTGTGCAAAATCGGTTTCGAGCTCCCATCGTTGGC 510  
QY 578 TCAGACACAGGGGCAAGGAACTCTGAGAGAGCTGTGGAGATGGAACAGCCACAAGAA 637  
DB 511 TCAGACACAGGGGCAAGGAACTCTGAGAGAGCTGTGGAGATGGAACAGCCACAAGAA 570  
QY 638 AACAGGGGAAAGCTTTTCAAGATGCTGTGACCAATTTTGAACCTCTGGAACCAAAAC 697  
DB 571 AACAGGGGAAAGCTTTTCAAGATGCTGTGACCAATTTTGAACCTCTGGAACCAAAAC 630  
QY 698 TCCAGTTCAGTGAAGAACTCTTCAAGATGTAAGCAAGTCCATATGATGCTGAGACAA 757  
DB 631 TCCAGTTCAGTGAAGAACTCTTCAAGATGTAAGCAAGTCCATATGATGCTGAGACAA 690  
QY 758 GTGCTAACAGGAATGATGATGAGGACCAAGTCCAGTCTGTGACATGAGGAGCTTT 817  
DB 691 GTGCTAACAGGAATGATGATGAGGACCAAGTCCAGTCTGTGACATGAGGAGCTTT 750  
QY 818 GTGCAAGCCACCCCTCTCCCAATTCATCTGCAAGTCTGAGCTGAGTGAACAT 877  
DB 751 GTGCAAGCCACCCCTCTCCCAATTCATCTGCAAGTCTGAGCTGAGTGAACAT 810  
QY 878 GTGCTGAAGATTTGTGTGAGACCTGAGAGCCACCCAGTGGGCTTAAGGTAAGCCAC 937  
DB 811 GTGCTGAAGATTTGTGTGAGACCTGAGAGCCACCCAGTGGGCTTAAGGTAAGCCAC 870  
QY 938 CAGTCCCAATGAGACTCACTGTGTGTGACCCAGAGACAGATTAAGCACTTGTCTAAGA 997  
DB 871 CAGTCCCAATGAGACTCACTGTGTGTGACCCAGAGACAGATTAAGCACTTGTCTAAGA 930  
QY 998 GGGGCTCTGCTTGAAGCTCATTAATCTTTTGTGTGACATTGAGACTGATGAGAGAT 1057  
DB 931 GGGGCTCTGCTTGAAGCTCATTAATCTTTTGTGTGACATTGAGACTGATGAGAGAG 990  
QY 1058 GGTGTGACAGATGATGTAGCTATTTTCAATTTAGATGAGCACTTCTTAAGTTAA 1117  
DB 991 GGGGAGGCTCCACCTTAAGCTCAACCAATTTTCAATTTAAAGGGGAGCTTCCAAATTAAG 1050  
QY 1118 GTTTATATGTTTGGGCAATA-TTTTGCTTAAGATATTTTAACTTTTATAC 1176

Db 1051 TTTATATATGTTTGGCAAAATTTTGTCTTAAATAATTTTAACTTTTAA 1110

Qy 1177 TTTGATTTTTCAGCTATTTTCTTAAAGATATATTTTCT 1219

Db 1111 CTTAAATTTTTCGCTATTTTCTTAAATAATTTT 1153

RESULT 4  
AV000968  
LOCUS  
DEFINITION  
AV000968 Mus musculus C57BL/6J kidney Mus musculus cDNA clone  
0610010F24, mRNA sequence.  
AV000968  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AV000968 989 bp mRNA linear EST 05-SEP-2001  
0610010F24, mRNA sequence.  
AV000968  
EST.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 989)  
Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T.,  
Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J.,  
Kono, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K.,  
Ono, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,  
Takeuchi, Y., Tanaka, T., Toyu, T., Muramatsu, M., and Hayashizaki, Y.  
Riken Mouse ESTs (Arakawa, T., et al. 2001)  
Unpublished (2001)  
On May 11, 1999 this sequence version replaced gi:4777818.  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsr.riken.go.jp,  
url:http://genome.gsc.riken.go.jp/  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
Matsubara, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A.  
and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P.,  
Sugahara, Y. and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Yamanaka, I., Kiyosawa, H., Kondo, S., Saito, T., Shinagawa, A.,  
Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K.,  
Arakawa, T., Ishii, Y. and Hayashizaki, Y.  
Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct.  
Funct. Genomics 2 pre, L72-L86 (2001)  
Please visit our web site (http://genome.gsc.riken.go.jp/) for  
further details.  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Location/Qualifiers  
1. 989  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"

Query Match 40.8%; Score 849.2; DB 9; Length 989;  
Best Local Similarity 92.9%; Pred. No. 5e-159;  
Matches 890; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

ORIGIN

/db\_xref="taxon:10090"  
/clone="0610010F24"  
/issue\_type="kidney"  
/clone\_lib="Mus musculus C57BL/6J kidney"

Qy 1103 ACTTCAAAATTAAGTTTATATGTTTGGCAATATTTGTCTTAAGATATTTT 1162

Db 31 ATTTCATTGAAGGAGGACTTAAATTATATATATGTTGCGATTATTTGCTAAGA 90

Qy 1163 TAAACTTTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAA 1222

Db 91 TATTTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAA 150

Qy 1223 AACATCCTGCTGCTCACTCACTTAAACATTTTAAACCTTAAACATTTTAAACATTT 1282

Db 151 AACACCTGCTGCTCACTCACTTAAACATTTTAAACATTTTAAACATTTTAAACATTT 210

Qy 1283 TAAAGTTTAAATAGAAATCTTTTGTGTTAGTCTCTACCTCCCAAGCAACTGT 1342

Db 211 TAAAGTTTAAATAGAAATCTTTTGTGTTAGTCTCTACCTCCCAAGCAACTGT 270

Qy 1343 AAATGTAGCCGCGCGGCTTTTACATAGAGGCTCCAGTAGGTCTTACATTTAGTAG 1402

Db 271 AAATGTAGCCGCGCGGCTTTTACATAGAGGCTCCAGTAGGTCTTACATTTAGTAG 330

Qy 1403 CTTGAAAACATGACAGCTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAG 1462

Db 331 CTTGAAAACATGACAGCTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAG 390

Qy 1463 AGCTCTTCTCTAGCCGCTGTGCGAGATGAGTCTTATTTATTTATTTATTTATTT 1522

Db 391 AGCTCTTCTCTAGCCGCTGTGCGAGATGAGTCTTATTTATTTATTTATTTATTTATTT 450

Qy 1523 CCACGTAAAGCTTAAAGTCTTACTCTGTAATCCCAACACAGTCTTCAAGGACTGCTG 1582

Db 451 CCACGTAAAGCTTAAAGTCTTACTCTGTAATCCCAACACAGTCTTCAAGGACTGCTG 510

Qy 1583 TGAGGAGAGCTTATGAGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1642

Db 511 TGAGGAGAGCTTATGAGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 570

Qy 1643 ACATGTGACATGAGACATGACATGACATGACATGACATGACATGACATGACATGACAT 1702

Db 571 ACATGTGACATGAGACATGACATGACATGACATGACATGACATGACATGACATGACATG 630

Qy 1703 TGTGTAGGTGAGACATGACATGACATGACATGACATGACATGACATGACATGACATG 1762

Db 631 TGTGTAGGTGAGACATGACATGACATGACATGACATGACATGACATGACATGACATG 690

Qy 1763 TGTGTGAGGAGGAGAGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGG 1822

Db 691 TGTGTGAGGAGGAGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAG 750

Qy 1823 TTTAAAGTTTAAACATTCGACATTCATATATTTAGGCTCTCCACCTGGGGGTGGCC 1882

Db 751 TTTAAAGTTTAAACATTCGACATTCATATATTTAGGCTCTCCACCTGGGGGTGGCC 810

Qy 1883 CCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1942

Db 811 CCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 870

Qy 1943 TGTATCTTTTGTGTTCTTGTGCTTTTGTGCTTTTGTGCTTTTGTGCTTTTGTGCTTTT 2002

Db 871 TGTATCTTTTGTGTTCTTGTGCTTTTGTGCTTTTGTGCTTTTGTGCTTTTGTGCTTTT 930

Qy 2003 ATTTTATATGCTTCAATATCAATATTAACCTCACTTATTTAAAGATATCAAG 2060

Db 931 ATTTTATATGCTTCAATATCAATATTAACCTCACTTATTTAAAGATATCAAG 988

RESULT 5  
LOCUS BU054465  
DEFINITION BU054465 836 bp mRNA linear EST 26-AUG-2002  
UI-M-FD0-bz3-e-04-0-UI.r1 NIH\_BMAP\_FD0 Mus musculus cDNA clone  
IMAGE:6404331 5', mRNA sequence.  
ACCESSION BU054465  
VERSION BU054465.1 GI:22494529  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 836)  
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Dr. James Lin, University of Iowa  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)  
Seq primer: PYX-5.  
FEATURES  
source  
Location/Qualifiers  
1..836  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6404331"  
/tissue\_type="whole brain"  
/dev\_stage="embryo 12.5 dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_1lb="NIH\_BMAP\_FD0"  
/note="Organ: brain; Vector: PYX-Asc; Site 1: EcoR I;  
Site 2: Not I; The library was constructed according to  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured mRNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with an  
oligo-dT primer containing a Not I site. Double stranded  
cDNA was size selected according to mRNA size fraction,  
ligated with EcoR I adaptor, digested with Not I, and then  
cloned directionally into PYX-Asc vector. The library tag  
sequence located between the Not I site and the polyA  
tail, is TGAGAGAGCC. This library was created for the  
University of Iowa Mouse Brain Molecular Anatomy Project  
(BMAP). 'Gene discovery in the developing Mouse Nervous  
System', supported by National Institutes of Mental Health  
(NIMH), Hemin Chin, Ph.D., program coordinator."

## ORIGIN

Query Match 38.9%; Score 810.4; DB 13; Length 836;  
Best local Similarity 99.5%; Pred. No. 3.2e-151;  
Matches 833; Conservative 0; Mismatches 2; Indels 2; Gaps 2;  
QY 252 GCACGTGCTCTTCTTCTAGCCCTGAGCGACAGTCACTCTT-GGACATGTCCTTTGTCAAG 310  
DB 1 GCACTGTCCCTTCTTCTTATAGCCCTGACGACAGTCACTCTTGTGACATGTCTTGTCAAG 60  
QY 311 CTCAGCTCTGTACATGCTAGTGTGAGCCCAATCTGCGCTCGGTCCATGCGCAAC 370  
DB 61 CTCACGCTCTGTACATGCTAGTGTGAGCCCAATCTTGCCTGCTGCTTCATGTGCAAC 120  
QY 371 ACAAGTCCGAGATCCAGAGAAATGAGCCAGATGTGTGTGTCATGGATGGACCC 430  
DB 121 ACAGTCCGAGATCCAGAGAAATGAGCCAGATGTGTGTGTCATGGATGGACCC 180

QY 431 CAGATGTAGATCGGGACCAAGTTGAACGCTGTGTGCCAGAGATCCTGTGTCTTACA 490  
DB 181 CAGATGTAGATCGGGACCAAGTTGAACGCTGTGTGCCAGAGATCCTGTGTCTTACA 240  
QY 491 GTGAGGGAGCTGAGAGAGAGACACCTGTCTCTTGAATCTGAGAGATGCTCCCTTGCMAAC 550  
DB 241 GTGAGGGAGCTGAGAGAGAGACACCTGTCTCTTGAATCTGAGAGATGCTCCCTTGCMAAC 300  
QY 551 TCGGTTTCCGAGCTCCCATGCTTGGCTCAGACACAGGGCAAGAAAGAACTCTCAGAGACG 610  
DB 301 TCGGTTTCCGAGCTCCCATGCTTGGCTCAGACACAGGGCAAGAAAGAACTCTCAGAGACG 360  
QY 611 CTCGCGGAGATGAGACAGCCCAAGAAACAGGGGAAAGCTTTCAGAGATGACAGTGGACAG 670  
DB 361 CTCGCGGAGATGAGACAGCCCAAGAAACAGGGGAAAGCTTTCAGAGATGACAGTGGACAG 420  
QY 671 ATATTGAGACCTGTGAGAACAAACAACTCCAGTTCAAGTGAAGAACTCTTCTGAGATGTG 730  
DB 421 ATATTGAGACCTGTGAGAACAAACAACTCCAGTTCAAGTGAAGAACTCTTCTGAGATGTG 479  
QY 731 GACAGCTCTTACTATGACTGTGACACAGTGTCTTAAAGAAATGATGAGTGGACCAAGTCC 790  
DB 480 GACAGCTCTTACTATGACTGTGACACAGTGTCTTAAAGAAATGATGAGTGGACCAAGTCC 539  
QY 791 AGTCTGTGCAATGCGCTGAGGGCTTGTCTGACAGCAACCCCTCTCCAGTTCCACTTGC 850  
DB 540 AGTCTGTGCAATGCGCTGAGGGCTTGTCTGACAGCAACCCCTCTCCAGTTCCACTTGC 599  
QY 851 AAGTGTGACCTGTGCTGAGCTGACACATGTGTGTAGATTTGTGTGAGACCTGAGAGGCC 910  
DB 600 AAGTGTGACCTGTGCTGAGCTGACACATGTGTGTAGATTTGTGTGAGACCTGAGAGGCC 659  
QY 911 ACCCCAGTGGGCTTAAAGGTGAGGCCACCACTCCCATGTGAGCTGACGTGTGTGAGACC 970  
DB 660 ACCCCAGTGGGCTTAAAGGTGAGGCCACCACTCCCATGTGAGCTGACGTGTGTGAGACC 719  
QY 971 AGAGACAGATTAAGCACTGTGCTTAAAGAGGGGCTCTGCTTGTAGCTCATTCCTTTTG 1030  
DB 720 AGAGACAGATTAAGCACTGTGCTTAAAGAGGGGCTCTGCTTGTAGCTCATTCCTTTTG 779  
QY 1031 TGTGACATTGGAATCACTGTGAGAGATGTGTGTCAAGCTTATGTCTTATTTT 1087  
DB 780 TGTGACATTGGAATCACTGTGAGAGATGTGTGTGTCAAGCTTATGTCTTATTTT 836

RESULT 6  
LOCUS CB183964  
DEFINITION CB183964 919 bp mRNA linear EST 31-JAN-2003  
AGENCOURT\_11380446 NIH\_MGC\_164 Mus musculus cDNA clone  
IMAGE:30245029 5', mRNA sequence.  
ACCESSION CB183964  
VERSION CB183964.1 GI:28183220  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 919)  
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Dr. David Rowe and Dr. Mina  
cDNA Library Preparation: Invitrogen Corp  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
Plate: NDAM0318 row: n column: 14  
High quality sequence stop: 674.



## FEATURES

source

Location/Qualifiers

1..919

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="IMAGE:30245029"

/lab\_host="DH10B (phage-resistant)"

/note="Vector: pCMV-SPORT6.1; Site\_1: EcoRV; Site\_2: NotI; Non-normalized full-length enriched library from pooled mouse embryonic limb, maxilla and mandible, day 10.5 and 11.5 (size selected for the 0.5-1 kb fragments) cloned directionally, priming method: Oligo-dT. cDNA enrichment: &gt;1k bp, Average insert size 1.8k bp. Priming sequence: 5'-GACTAGTCTAGTGGGAGGCGGCCCC(T)3'. Tissue contributed by, David Rowe. Library constructed by Resgen, Invitrogen Corp."

## ORIGIN

Query Match 38.9%; Score 809.8; DB 14; Length 919;  
 Best Local Similarity 95.5%; Pred. No. 4e-151;  
 Matches 878; Conservative 0; Mismatches 32; Indels 9; Gaps 4;

```

QY 1064 TCACAGCTATGCTAGTCTATTTTCAATTAGTAGTGAACCTTTCTAAATTAAGTTTGA 1123
DB 1 TCACAGCTATGCTAGTCTATTTTCAATTAGTAGTGAACCTTTCTAAATTAAGTTTGA 60
QY 1124 TATGTTTTGGGCAATATTTTGTCTAGATATATTTTAAACTTTTAACTTTTAACTTTGAT 1183
DB 61 TATGTTTTGGGCAATATTTTGTCTAGATATATTTTAACTTTTAACTTTTAACTTTGAT 120
QY 1184 TTTTTCAGCTATTTTCTTAAAGTATATTTTCTACAAACCTCTGCTGCTACAT 1243
DB 121 TTTTTCAGCTATTTTCTTAAAGTATATTTTCTACAAACCTCTGCTGCTACAT 180
QY 1244 AGAACAATTTATTAACCTTAATAGATGTTGTTGCTCATTTTAAAGTTTAAATGAAGAACT 1303
DB 181 AGAACAATTTATTAACCTTAATAGATGTTGTTGCTCATTTTAAAGTTTAAATGAAGAACT 240
QY 1304 TCTTTTGTACTAGTCTCTACACTCCCAAGGCACTGTAATGTAGCGGCGGGGTGT 1363
DB 241 TCTTTTGTACTAGTCTCTACACTCCCAAGGCACTGTAATGTAGCGGCGGGGTGT 300
QY 1364 TACATGAGAGGCTCCAGTATGTTCTACATTTCTAGTAGAGCTTGAAA-AGAACATGCA 1421
DB 301 TACATGAGAGGCTCCAGTATGTTCTACATTTCTAGTAGAGCTTGAAAAGAACATGCA 360
QY 1422 AGCTCCAGTCCCGCTCAGTGGGTCTGCTGGCGGATCGGAGCTCTTTCTAGCCCG 1481
DB 361 AGCTCCAGTCCCGCTCAGTGGGTCTGCTGGCGGATCGGAGCTCTTTCTAGCCCG 420
QY 1482 TGTGAGAGTGGCTTTATTTATGCTATTTATATGTAATGCACTGAAAGCTTAAAGTCT 1541
DB 421 TGTGAGAGTGGCTTTATTTATGCTATTTATATGTAATGCACTGAAAGCTTAAAGTCT 480
QY 1542 TACTCTGGAATATCCCAACCAAGTCTTCAAGGAGCTGCTGTAGGAGCACTGCTTATGCA 1601
DB 481 TACTCTGGAATATCCCAACCAAGTCTTCAAGGAGCTGCTGTAGGAGCACTGCTTATGCA 540
QY 1602 GGTCTTG---TCTTGGGCATCACTGTCTGGTTCCAGGCCAGCACTGTAACATGAG 1657
DB 541 GGTCTTGCTTGTCTTGGGCATCACTGTCTGGTTCCAGGCCAGCACTGTAACATGAG 600
QY 1658 ACATGACATGCCGAACCAACCAAGCACTGCTCATGCTCAAGTGTGTAGTGAAGAC 1717
DB 601 ACATGACATGCCGAACCAACCAAGCACTGCTCATGCTCAAGTGTGTAGTGAAGAC 660
QY 1718 CACTGGCTCCAGGAGGCTGTGCTCAGAGAGGTTGTGAGTCTTCAAGTGTGCTGGGAGAC 1777
DB 661 CACTGGCTCCAGGAGGCTGTGCTCAGAGAGGTTGTGAGTCTTCAAGTGTGCTGGGAGAC 720
QY 1778 GACGGTGAAGCTGTGCTTGTCTTTAAATGAGTGTGAGCTTTTAAAGTTTAAAGAC 1837
DB 1778 GACGGTGAAGCTGTGCTTGTCTTTAAATGAGTGTGAGCTTTTAAAGTTTAAAGAC 1837

```

DB

721 GACGGTGAAGCTGTGCTTGTCTTTT-AAATGAGTGTGAGTGTTTAAAGTTTAAAGAC 779

QY

1838 AATCCGACATCATATATATTTAGGGCTCTCCACCTGGGGGTGGCCCTTATGCTGCT 1897

DB

780 AATCCGACATCATATATTTAGGGCTCTCTCCACCTGGGGGTGGCCCTTATGCTGCT 839

QY

1898 TGGATCTCAAGTCTTGTGACTCGG--CACTGTCAAGCTCCACCCCATGATCTTTTGG 1955

DB

840 TGGATCTCAAGTCTTGTGACTCGGCACTGTGTCAAGCTTACCCCAAGACCCCTTTT 899

QY

1956 TTCTCTGTGCTTTT 1974

DB

900 GGTTCCTCCCTGGGCTTT 918

## RESULT 7

BQ950078

LOCUS

DEFINITION

BQ950078 927 bp mRNA linear EST 21-AUG-2002

AGENCOURT\_888587 NIH\_MGC\_129 Mus musculus cDNA clone IMAGE:6311335

5', mRNA sequence.

ACCESSION

BQ950078

VERSION

BQ950078.1 GI:22365556

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: g3apb@remail.nih.gov

Tissue Procurement: Susan L. Sullivan, PhD.

CDNA Library Preparation: Resgen, Invitrogen Corp

DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://image.lnl.gov

Plate: LLM13735 row: b column: 08

High quality sequence stop: 608.

## FEATURES

source

Location/Qualifiers

1..927

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="IMAGE:6311335"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH MGC 129"

/note="Organ: olfactory epithelium; Vector: pCMV-SPORT6.1; Site\_1: EcoRV; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.2 kb. Constructed by Resgen, Invitrogen Corp. Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 38.3%; Score 798; DB 13; Length 927;  
 Best Local Similarity 96.6%; Pred. No. 9.1e-149;  
 Matches 858; Conservative 0; Mismatches 25; Indels 5; Gaps 4;

```

QY 182 CTGACAGAGATGTTGCTAGAGGCTGAAGAGAAATATGTTGACCAAGAGAAAGAGTA 241
DB 22 CCGACAGAGATGTTGCTAGAGGCTGAAGAGAAATATGTTGACCAAGAGAAAGAGTA 81
QY 242 GAGGTTTTGGCACTGTCTCTTATAGCTTCAAGAGACAGTCACTCTGACATGTCC 301
DB 82 GAGGTTTTGGCACTGTCTCTTATAGCTTCAAGAGACAGTCACTCTGACATGTCC 141
QY 302 CTGTCAAGCTCAAGCTCTGTCAAGTGTGAGAGCCCAATCTTGGCGCTGGTCTTC 361
DB 142 CTGTCAAGCTCAAGCTCTGTCAAGTGTGAGAGCCCAATCTTGGCGCTGGTCTTC 201

```



QY 362 ATGCGCAACAGCTCCGCGAGATCCAGAGAGAAATGACCGAGATGTGTGTGGCAATGGG 421  
Db 202 ATGCGCAACAGCTCCGCGAGATCCAGAGAGAAATGACCGAGATGTGTGTGGCAATGGG 261  
QY 422 ATGCGCAACAGCTCCGCGAGATCCAGAGAGAAATGACCGAGATGTGTGTGGCAATGGG 481  
Db 262 ATGCGCAACAGCTCCGCGAGATCCAGAGAGAAATGACCGAGATGTGTGTGGCAATGGG 321  
QY 482 TGTGCTGACGTGAGGGGAGCTGAGAGAGACACCCCTGCTCTGAACCTGGAAGATGCTCCC 541  
Db 322 TGTGCTGACGTGAGGGGAGCTGAGAGAGACACCCCTGCTCTGAACCTGGAAGATGCTCCC 381  
QY 542 TTGCAAAACTCGGTTTCCGAGCTCCCATGTTGCTCAGACACCGAGGCAAAAGAAACCTT 601  
Db 382 TTGCAAAACTCGGTTTCCGAGCTCCCATGTTGCTCAGACACCGAGGCAAAAGAAACCTT 441  
QY 602 CAGAGCAGCTCTGGGAGATGAGACAGCCCAAGAAAGAGGGGAAAGCTTTTCAAGAGTCA 661  
Db 442 CAGAGCAGCTCTGGGAGATGAGACAGCCCAAGAAAGAGGGGAAAGCTTTTCAAGAGTCA 501  
QY 662 CTGGACCAATATTTTGAAGACCTGGAGAAACAAATCCAGTTCAAGTGAAGAACTTTC 721  
Db 502 CTGGACCAATATTTTGAAGACCTGGAGAAACAAATCCAGTTCAAGTGAAGAACTTTC 561  
QY 722 TCAGATGTGAGACAGCTCTCTACTATGACCTGAGACAGTGTAAACAGGAATGATGAGTGG 781  
Db 562 TCAGATGTGAGACAGCTCTCTACTATGACCTGAGACAGTGTAAACAGGAATGATGAGTGG 621  
QY 782 ACCAAGTCCAGTCTGTGCAATGAGCTTTGAGGGCTTTGCTGAGACCAACCTCTCTCCAGT 841  
Db 622 ACCAAGTCCAGTCTGTGCAATGAGCTTTGAGGGCTTTGCTGAGACCAACCTCTCTCCAGT 681  
QY 842 TCACCTTTGCAAGCTCTGACCTGCTGAGCTGAGACCAATGTGTAGAGATTCTGTGTGAGAGC 901  
Db 682 TCACCTTTGCAAGCTCTGACCTGCTGAGCTGAGACCAATGTGTAGAGATTCTGTGTGAGAGC 741  
QY 902 TGAGAGGCCACCCCAAGTGGG-CTAAGGGTGAAGGCCACAGTCCCAATGAGCTCAAGTGT 960  
Db 742 TGAGAGGCCACCCCAAGTGGGCTTAAGGGTGAAGGCCACAGTCCCAATGAGCTCAAGTGT 801  
QY 961 GTTGTGACCCAGAGACAGATAGCACTT-GTCTTAAGAGGGGCTCTGG-CTCTTAAGCT 1017  
Db 802 GGTGTGACCCAGAGACAGATAGCACTTGTCTTAAGAGGGGCTCTGGGCTCTCTGAGCT 861  
QY 1018 CATT-ATCTTTTGTGTGACATGTGACCTCACTGTGAGAGATGTGTGT 1064  
Db 862 CATTATCTTTTGTGTGAGATGTGAGACCTCTGGGGAAGAGAGGT 909

RESULT 8  
BUI48336 920 bp mRNA linear EST 03-SEP-2002  
LOCUS BUI48336  
DEFINITION AGENCOURT\_8753844 NIH\_MGC\_130 Mus musculus cDNA clone IMAGE:6393860  
5', mRNA Sequence.  
ACCESSION BUI48336  
VERSION BUI48336.1 GI:22661868  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 920)  
NIH-MGC <http://mgi.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [csabers-remail.nih.gov](mailto:csabers-remail.nih.gov)  
Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,  
Ph.D.  
cDNA Library Preparation: ResGen, Invitrogen Corp  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be

Found through the I.M.A.G.E. Consortium/LIML at:  
<http://image.liml.gov>  
Plate: LIML3885 row: p column: 21  
High quality sequence stop: 740.  
Location/Qualifiers  
1..920  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6393860"  
/lab\_host="DH10B (phage-resistant)"  
/note="Orgam: olocysts; Vector: PCMV-SPORT6.1; Site: 1;  
Ecory: Site 2: Not; Cloned unidirectionally. Primer:  
Oligo dT. Average insert size 1.95 kb. Constructed by  
ResGen, Invitrogen Corp. Note: this is a NIH\_MGC Library."

ORIGIN  
Query Match 38.0%; Score 790.6; DB 13; Length 920;  
Best Local Similarity 98.7%; Pred. No. 2,7e-147;  
Matches 815; Conservative 0; Mismatches 9; Indels 2; Gaps 2;  
QY 182 CTGGACAGAGTGTCTGAGAGGCTGAAGAGAAATATGTGACCAAGAGAGAGTA 241  
Db 46 CCGACAGAGATGTTTCTGAGAGGCTGAAGAGAAATATGTGACCAAGAGAGAGTA 105  
QY 242 GAGGTTTGGCACTGCTCCCTTCTTAAGCTTGAAGGACAGTCACTCTCTGACATGTCC 301  
Db 106 GAGGTTTGGCACTGCTCCCTTCTTAAGCTTGAAGGACAGTCACTCTCTGACATGTCC 165  
QY 302 CTGTGCAAGTCCAGCTCTGTCAATGCTAGTGAAGCCCAATCTGCGGCTGCTCTC 361  
Db 166 CTGTGCAAGTCCAGCTCTGTCAATGCTAGTGAAGCCCAATCTGCGGCTGCTCTC 225  
QY 362 ATGCGCAACAGCTCCGCGAGATCCAGAGAGAAATGACCGAGATGTGTGTGGCAATGGG 421  
Db 226 ATGCGCAACAGCTCCGCGAGATCCAGAGAGAAATGACCGAGATGTGTGTGGCAATGGG 285  
QY 422 ATGCGCAACAGCTCCGCGAGATCCAGAGAGAAATGACCGAGATGTGTGTGGCAATGGG 481  
Db 286 ATGCGCAACAGCTCCGCGAGATCCAGAGAGAAATGACCGAGATGTGTGTGGCAATGGG 345  
QY 482 TGTGCTGACGTGAGGGGAGCTGAGAGAGACACCTGCTCTGAACTGGAAGATGCTCCC 541  
Db 346 TGTGCTGACGTGAGGGGAGCTGAGAGAGACACCTGCTCTGAACTGGAAGATGCTCCC 405  
QY 542 TTGCAAAACTCGGTTTCCGAGCTCCCATGTTGCTCAGACACCGAGGCAAAAGAAACCTT 601  
Db 406 TTGCAAAACTCGGTTTCCGAGCTCCCATGTTGCTCAGACACCGAGGCAAAAGAAACCTT 465  
QY 602 CAGAGCAGCTCTGGGAGATGAGACAGCCCAAGAAAGAGGGGAAAGCTTTTCAAGAGTCA 661  
Db 466 CAGAGCAGCTCTGGGAGATGAGACAGCCCAAGAAAGAGGGGAAAGCTTTTCAAGAGTCA 525  
QY 662 CTGGACCAATATTTTGAAGACCTGGAGAAACAAATCCAGTTCAAGTGAAGAACTTTC 721  
Db 526 CTGGACCAATATTTTGAAGACCTGGAGAAACAAATCCAGTTCAAGTGAAGAACTTTC 585  
QY 722 TCAGATGTGAGACAGCTCTCTACTATGACCTGAGACAGTGTAAACAGGAATGATGAGTGG 781  
Db 586 TCAGATGTGAGACAGCTCTCTACTATGACCTGAGACAGTGTAAACAGGAATGATGAGTGG 645  
QY 782 ACCAAGTCCAGTCTGTGCAATGAGCTTTGAGGGCTTTGCTGAGACCAACCTCTCTCCAGT 841  
Db 646 ACCAAGTCCAGTCTGTGCAATGAGCTTTGAGGGCTTTGCTGAGACCAACCTCTCTCCAGT 705  
QY 842 TCACCTTTGCAAGCTCTGACCTGCTGAGCTGAGACCAATGTGTAGAGATTCTGTGTGAGAGC 901  
Db 706 TCACCTTTGCAAGCTCTGACCTGCTGAGCTGAGACCAATGTGTAGAGATTCTGTGTGAGAGC 765  
QY 902 TGAGAGGCCACCCCAAGTGGGCTTAAGGGTGAAGGCCACAGTCCCAATG-AGTCAAGTGT 960  
Db 766 TGAGAGGCCACCCCAAGTGGGCTTAAGGGTGAAGGCCACAGTCCCAATGAGAGTCAAGTGT 825

QY 961 GTTGT-GACCCAGAGACAGATTAAGCACTTGTCTTAAGAGGGCTCT 1005  
DB 826 GGTGTGACCCAGAGACAGATTAAGCACTTGTCTTAAGAGGGAGCT 871

RESULT 9  
BUI51870  
LOCUS  
DEFINITION  
AGENCOURT\_8754199 NIH\_MGC\_130 Mus musculus cDNA clone IMAGE:6332467  
5', mRNA sequence.  
BUI51870  
VERSION  
BUI51870.1 GI:22665402  
KEYWORDS  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 926)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapdb-remail.nih.gov](mailto:cgapdb-remail.nih.gov)  
Tissue Procurement: Mark Macconochie, Ph.D. and Nancy L. Freeman,  
Ph.D.  
cDNA Library Preparation: ResGen, Invitrogen Corp  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
Plate: L1AM13790 row: b column: 20  
High quality sequence stop: 660.  
Location/Qualifiers  
1..926  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6332467"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: oocytes; Vector: pCMV-Sport6.1, Site\_1:  
EcoRV, Site\_2: NotI; Cloned unidirectionally. Primer:  
Oligo dt. Average insert size 1.95 kb. Constructed by  
ResGen, Invitrogen Corp. Note: this is a NIH\_MGC Library."

ORIGIN  
Query Match 37.7%; Score 785.8; DB 13; Length 926;  
Best Local Similarity 95.8%; Pired. No. 2,4e-146;  
Matches 883; Conservative 0; Mismatches 25; Indels 14; Gaps 7;

QY 134 TCCGGGTGCCCGAAGAGGCTGGCGTGGTCTGCTTGCAGCCGCCGCTGGAGAGGAT 193  
DB 6 TCCGGGTGCCCGAAGAG--TGGCGGGGTCTGCTTGCAGC-----GCCGAGCAGATG 57

QY 194 TTTGCTAGAGGGCTGAAGAAATATGTGTACCAAGAAAGAGATAGAGGATTTGGC 253  
DB 58 TTTGCTAGAGGGCTGAAGAAATATGTGTACCAAGAAAGAGATAGAGGATTTGGC 117

QY 254 ACTGTCCCTTCATATGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 313  
DB 118 ACTGTCCCTTCATATGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 177

QY 314 CAGCTCTGTACATGTGTAG 373  
DB 178 CAGCTCTGTACATGTGTAG 237

QY 374 GTCCGCGAGATCCAG 433  
DB 238 GTCCGCGAGATCCAG 297

QY 434 AATGTAGATCGGGACCAAGTTGAAGCGCTGGTGTCCACAGAGATCCTGTGTCTGATG 493

DB 298 AATGTAGATCGGGACCAAGTTGAAGCGCTGGTGTCCACAGAGATCCTGTGTCTGATG 357  
QY 494 AGGGAGCTGAG 553  
DB 358 AGGGAGCTGAG 417  
QY 554 GTTTCGAGAGCTCCCATGCTTGGCTGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAG 613  
DB 418 GTTTCGAGAGCTCCCATGCTTGGCTGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAG 477

QY 614 TGGGAGATGAGACAGCCCAAG 673  
DB 478 TGGGAGATGAGACAGCCCAAG 537

QY 674 TTTGAGACCTTGAG 733  
DB 538 TTTGAGACCTTGAG 597

QY 734 AGCTCTACTATGACCTGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 793  
DB 598 AGCTCTACTATGACCTGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 657

QY 794 CTCTGCAATGAGCTTGAG 853  
DB 658 CTCTGCAATGAGCTTGAG 717

QY 854 TCTGACCTGAGCTGAG 912  
DB 718 TCTGACCTGAGCTGAG 777

QY 913 CCCAGT-GGGCTAAGAGGAG 971  
DB 778 CCCAGTGGGGCTAAGAGGAG 837

QY 972 GAGACAGATTAAGCACTTGTCTTAAGA--GGGGCTGTGGCTTTAGAGTCATT-ATCCTTT 1028  
DB 838 GAGACAGATTAAGCACTTGTCTTAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 897

QY 1029 TGTGT-GACATTGAGACTCAGTG 1049  
DB 898 TGTGTGACATTGAGACTCCTTG 919

RESULT 10  
BBI09465  
LOCUS  
DEFINITION  
BBI09465 RIKEN full-length enriched, adult male kidney Mus musculus  
cDNA clone 0610010F24 5', mRNA sequence.  
BBI09465  
VERSION  
BBI09465.1 GI:15390279  
KEYWORDS  
SOURCE  
EST.  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 994)

REFERENCE  
AUTHORS  
Arai, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J.,  
Kono, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K.,  
Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shitaki, T.,  
Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takehashi, F.,  
Takeda, Y., Tanaka, T., Toyota, T., Mizumatsu, M., and Hayashizaki, Y.  
RIKEN Mouse ESTs (Arai, et al. 2001)  
Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: genome-res@sc.riken.go.jp,  
 URL: http://genome.gsc.riken.go.jp/  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. *Genome Res.* 10 (10), 1617-1630 (2000)  
 wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
 Matsubara, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A.,  
 and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer. *Genome Res.*  
 10 (11), 1757-1771 (2000)  
 Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P.,  
 Sugahara, Y., and Hayashizaki, Y.  
 Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)  
 Yamanaka, I., Kiyosawa, H., Kondo, S., Saito, T., Shingawa, A.,  
 Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K.,  
 Akakawa, T., Ishii, Y., and Hayashizaki, Y.  
 Mapping of 19032 mouse cDNAs on mouse chromosomes. *J. Struct.*  
*Func. Genomics* 2 pre, L72-L86 (2001)  
 Please visit our web site (http://genome.gsc.riken.go.jp) for  
 further details.  
 e mouse tissues.

## FEATURES

## source

Location/Qualifiers  
 1. 994  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="0610010P24"  
 /sex="male"  
 /tissue\_type="kidney"  
 /dev\_stage="adult"  
 /lab\_host="SOLR"  
 /clone\_1ib="RIKEN full-length enriched, adult male kidney"  
 /note="Site\_1: XhoI; Site\_2: SctI; cDNA library was  
 prepared and sequenced in Mouse Genome Encyclopedia  
 Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in  
 RIKEN. Division of Experimental Animal Research in Riken  
 contributed to prepare mouse tissues. 1st strand cDNA was  
 primed with a primer [5'  
 GAGAGAGAGCGCGCGCACTGAGTTTCTTTTCTTTTNN 3'], cDNA was  
 prepared by using trehalose thermo-activated reverse  
 transcriptase and subsequently enriched for full-length by  
 cap-trapper. Second strand cDNA was prepared with the  
 primer adapter of sequence [5'  
 GAGAGAGAGAGATCCAGAGCTCAATTAATTAATTAACCCCCCCCC 3'].  
 cDNA was cleaved with XhoI and SctI."

## ORIGIN

Query Match 36.3%; Score 755.4; DB 10; Length 994;  
 Best Local Similarity 96.2%; Pred. No. 2.7e-140;  
 Matches 815; Conservative 0; Mismatches 23; Indels 9; Gaps 4;  
 Db 182 CTGACAGATGTTTGTCTAGAGGGCTGAGAGAAATATGTGACCCAGAGAGAGAGTA 241  
 Db 83 CCGAGACGATGTTTGTCTAGAGGGCTGAGAGAAATATGTGACCCAGAGAGAGAGTA 142  
 Db 242 GAGGGTTTGGCACTGTCCTTCTTATAGCTCTGAGGAGCAATCTCTGAGCATGTC 301  
 Db 143 GAGGGTTTGGCACTGTCCTTCTTATAGCTCTGAGGAGCAATCTCTGAGCATGTC 202  
 Db 302 CTGTCAAGCTCCAGCTCTGTGACATCTAGTGGAGCCCAATCTCTGCGTGGTCTC 361  
 Db 203 CTGTCAAGCTCCAGCTCTGTGACATCTAGTGGAGCCCAATCTCTGCGTGGTCTC 262  
 Db 362 ATGCCCAACACAGTCCGGCAGATCCAGAGAGAAATGAGCAGAGATGCTGTGTGAGCATGG 421  
 Db 263 ATGCCCAACACAGTCCGGCAGATCCAGAGAGAAATGAGCAGAGATGCTGTGTGAGCATGG 322

Qy 422 ATGCACCCCCAGATGATGATCGGGACACAGTGAACGCTGGTGTCCACAGATCTG 481  
 Db 323 ATGCACCCCCAGATGATGATCGGGACACAGTGAACGCTGGTGTCCACAGATCTG 382  
 Qy 482 TGTCTACAGTGAAGGGAGCTGAGAGAGACCTGCTCTTGAACCTGAAGATGCTCC 541  
 Db 383 TGTCTACAGTGAAGGGAGCTGAGAGAGACCTGCTCTTGAACCTGAAGATGCTCC 442  
 Qy 542 TTGCAGAACTCGGTTTCCGAGCTCCCATGTGTTGGCTCAGACACAGGGCAAGACCT 601  
 Db 443 TTGCAGAACTCGGTTTCCGAGCTCCCATGTGTTGGCTCAGACACAGGGCAAGACCT 502  
 Qy 602 CAGAGAGCTCTGGGAGATGAGCAGGCCCAAGAAAACAGGGGAAAGCTTTCAGAGTCA 661  
 Db 503 CAGAGAGCTCTGGGAGATGAGCAGGCCCAAGAAAACAGGGGAAAGCTTTCAGAGTCA 562  
 Qy 662 CTGACAGATATTTGAGACCTCTGAGAAACAAAACCTCAGTTCAGTGAAGAACTTTC 721  
 Db 563 CTGACAGATATTTGAGACCTCTGAGAAACAAAACCTCAGTTCAGTGAAGAACTTTC 622  
 Qy 722 TCGATGTGAGCAGCTCTTACTATGACCTGAGACACAGTGTCTAAGAGATGATGAGTGG 781  
 Db 623 TCGATGTGAGCAGCTCTTACTATGACCTGAGACACAGTGTCTAAGAGATGATGAGTGG 682  
 Qy 782 ACCAGTCCAGTCTCTGCAATGAGCTTGAAGGCTTGTGACAGCAACCTCTCCAGT 841  
 Db 683 ACCAGTCCAGTCTCTGCAATGAGCTTGAAGGCTTGTGACAGCAACCTCTCCAGT 742  
 Qy 842 TCCACTTGCAGATCTGACCTGCTGAGCTGAGCCATGTGTGATGATGATGAGTGAAC 901  
 Db 743 TCCACTTGCAGATCTGACCTGCTGAGCTGAGCCATGTGTGATGATGATGAGTGAAC 802  
 Qy 902 TGAAGGCCACCCCACTGAGGCTTAAGGCTGAGGCCACAGTCCCATGAGACTACCTGTG 961  
 Db 803 TGAAGGCCACCCCACTGAGGCTTAAGGCTGAGGCCACAGTCCCATGAGACTACCTGTG 859  
 Qy 962 TTGTTGCCACAGACAGATGACATGCTGCTTAAGAGGGGCTGCTGTGAGCTCAT 1021  
 Db 860 TTGTTGCCACAGATGACATGCTGCTTAAGAGGGGCTGCTGTGAGCTCAT 913  
 Qy 1022 ATCCTTT 1028  
 Db 914 ATCCTTT 920

RESULT 11  
 BO936878  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLNL4035 row: 1 column: 12  
 High quality sequence stop: 635.

Location/Qualifiers  
1. .967

ORIGIN

Query Match  
Best Local  
Matches 78

Y	215
5	1
Y	275

61  
335

395  
181

455  
241  
515

301  
575  
361

635  
421

481  
755

541  
815  
601

875  
661  
025

721

QY  
Db

RESULT 12  
BQ713776  
LOCUS  
DEFINITION

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

11

## FEATURES

Query Ma  
Best Loc  
Matches

QY  
Db  
Qv

QY

QY 2

Db 2

QY	3
Db	2
QY	4

Db 317 AATGTAGATCGGCGACCAAGTTGAAAGCGCTGGTGTCCACAGAGATCTGTGTCGTACAGTG 376  
 Qy 494 AGGGAGCTGAGAGAGACCCGCTCTGTAAGTGTGCTCCCTTGGCAAACTCG 553  
 Db 377 AGGGAGCTGAGAGAGACCCGCTCTGTAAGTGTGCTCCCTTGGCAAACTCG 436  
 Qy 554 GTTCCGAGCTCCCATCTGTTGGCTCAGACCAAGGCAAAAGAACCTTCAGAGAGCTC 613  
 Db 437 GTTCCGAGCTCCCATCTGTTGGCTCAGACCAAGGCAAAAGAACCTTCAGAGAGCTC 496  
 Qy 614 TGGAGATGAGACAGCCCAAGAAAGAGGGAAGCTTTGAGAGTCACTGAGCAAGATA 673  
 Db 497 TGGAGATGAGACAGCCCAAGAAAGAGGGAAGCTTTGAGAGTCACTGAGCAAGATA 556  
 Qy 674 TTTGAGACCTTGGAGAAACAAACTCCAGTTCACTGAGAGAACTTTCTCAGATGTGAC 733  
 Db 557 TTTGAGACCTTGGAGAAACAAACTCCAGTTCACTGAGAGAACTTTCTCAGATGTGAC 616  
 Qy 734 AGCTCTACTATGACCTGAGACAGTGTCTAAAGAGATGATGAGGAGCAAGTCCAGT 793  
 Db 617 AGCTCCGACTATGACCTGAGACAGTGTCTAAAGAGATGATGAGGAGCAAGTCCAGT 676  
 Qy 794 CTCTCAATGAGCTTGAAGGCTTTCTGAGAGCAAGCTCCCTCCAGTTCCACTTGAAG 853  
 Db 677 CTCTCAATGAGCTTGAAGGCTTTCTGAGAGCAAGCTCCCTCCAGTTCCACTTGAAG 736  
 Qy 854 TCTGACCTGCTGAGCTGAGACCATGTGTAGATTTCTGCTGAGACCTGAGAGGCTC 912  
 Db 737 TCTGACCTGCTGAGCTGAGACCATGTGTAGATTTCTGCTGAGAGCTGAGAGGCTC 796  
 Qy 913 CCCAGTGGGCTAAGGAGTGAAGGCA-CCAGTCCCATGAGAGTCACTGTGTGTG---GA 967  
 Db 797 CCCAGTGGGCTAAGGAGTGAAGGCAAGTCCCATGAGAGTCACTGTGTGTGTGAG 856  
 Qy 968 CCCAGAGACAGATTAAGC 984  
 Db 857 CCCAGAGACAGATTAAGC 873

RESULT 13  
 Bi61691  
 LOCUS 830 bp mRNA linear EST 12-SEP-2001  
 DEFINITION 60305632F1\_NCI\_CGAP\_Mam4 Mus musculus cDNA clone IMAGE:5351646,  
 mRNA sequence.  
 ACCESSION Bi61691  
 VERSION Bi61691.1 GI:15575927  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 830)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strauberg, Ph.D.  
 Email: cga@rs-remail.nih.gov  
 Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furch  
 Ph.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.lnl.gov  
 Plate: L1AM11893 row: 0 column: 07  
 High quality sequence stop: 797.  
 Location/Qualifiers  
 1..830  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="NMRI"  
 /db\_xref="taxon:10090"

/clone="IMAGE:5351646"  
 /issue\_type="tumor, gross tissue"  
 /dev\_stage="5 months"  
 /lab\_host="DH10B"  
 /clone\_lib="NCI\_CGAP\_Mam4"  
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;  
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
 Library constructed by Life Technologies. Investigators  
 providing samples: Lothar Hennighausen, Priscilla Furch,  
 NIH Reference for transgenic model: Li et al., Cell Growth  
 and Differentiation 7, 3-11 (1996)."

Query Match 35.6%; Score 741; DB 12; Length 830;  
 Best Local Similarity 98.7%; Pred. No. 2.1e-137;  
 Matches 755; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

Qy 182 CTGACACAGATGTTGCTAGAGGGCTGAGAGGAATATGATGACAGAGAGAGAGTA 241  
 Db 66 CCGAGACAGATGTTGCTAGAGGGCTGAGAGGAATATGATGACAGAGAGAGAGTA 125  
 Qy 242 GAGGTTTGGCACTGTCCCTTCTATAGCTCTGAGAGCAATCATCTCTGACATGTC 301  
 Db 126 GAGGTTTGGCACTGTCCCTTCTATAGCTCTGAGAGCAATCATCTCTGACATGTC 185  
 Qy 302 CTGTCAGAGCTGCAAGCTCTGCAATGCTGAGAGCCCAATCTGCGCTGCTGCTC 361  
 Db 186 CTGTCAGAGCTGCAAGCTCTGCAATGCTGAGAGCCCAATCTGCGCTGCTGCTC 245  
 Qy 362 ATGCGCAACACAGTCCGCGAGATCCAGAGGAATAGAGAGATGATGATGATGATG 421  
 Db 246 ATGCGCAACACAGTCCGCGAGATCCAGAGGAATAGAGAGATGATGATGATGATG 305  
 Qy 422 ATGCGCAACACAGTCCGCGAGATCCAGAGGAATAGAGAGATGATGATGATGATG 481  
 Db 306 ATGCGCAACACAGTCCGCGAGATCCAGAGGAATAGAGAGATGATGATGATGATG 365  
 Qy 482 TGTGCTACAGTGAAGGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 541  
 Db 366 TGTGCTACAGTGAAGGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 425  
 Qy 542 TTGCAAAACTCGGTTTCCAGAGCTCCCATGCTGTTGGCTCAGACCAAGGCAAAAGCCCT 601  
 Db 426 TTGCAAAACTCGGTTTCCAGAGCTCCCATGCTGTTGGCTCAGACCAAGGCAAAAGCCCT 485  
 Qy 602 CAGAGAGCTCTGAGAGATGAGACAGCCCAAGAAAGAGGGAAGCTTTCAGAGTCA 661  
 Db 486 CAGAGAGCTCTGAGAGATGAGACAGCCCAAGAAAGAGGGAAGCTTTCAGAGTCA 545  
 Qy 662 CTGAGACCAAGTATTTGAGAGAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAG 721  
 Db 546 CTGAGACCAAGTATTTGAGAGAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAG 605  
 Qy 722 TCAGATGAGAGAGCTCTGAGAGAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAG 781  
 Db 606 TCAGATGAGAGAGCTCTGAGAGAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAG 665  
 Qy 782 ACCAAGTCCAGTCTGAGAGAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 841  
 Db 666 ACCAAGTCCAGTCTGAGAGAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 725  
 Qy 842 TCCACTTGGCAAGTCTGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 901  
 Db 726 TCCACTTGGCAAGTCTGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 785  
 Qy 902 TGAGAGGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 946  
 Db 786 TGAGAGG-CACCAAGTGGGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 829

RESULT 14  
 B1153019  
 LOCUS B1153019 800 bp mRNA linear EST 05-JUL-2001

DEFINITION 602918211F1 NCI\_CGAP\_Lu29 Mus musculus cDNA clone IMAGE:5068418 5', mRNA sequence.  
 ACCESSION BI153019  
 VERSION BI153019.1 GI:14613020  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 800)  
 NIH-MGC http://mgi.nci.nih.gov/.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 CONTACT Robert Strausberg, Ph.D.  
 Email: cgsrbs-remail.nih.gov  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LIML at: http://image.llnl.gov  
 Plate: LIML1184 row: f column: 03  
 High quality sequence stop: 797.  
 Location/Qualifiers  
 1..800  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="Czech 11"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5068418"  
 /tissue\_type="spontaneous tumor, metastatic to mammary."  
 /cell\_line="DHI0B"  
 /lab\_host="DHI0B"  
 /clone\_1b="NCI CGAP Lu29"  
 /note="Organ: lung; Vector: pCMV-SPORT6; Site: 1: Salt; Site 2: Not; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

ORIGIN  
 Query Match 34.6%; Score 721.4; DB 12; Length 800;  
 Best Local Similarity 98.3%; Pred. No. 1.8e-133;  
 Matches 770; Conservative 0; Mismatches 8; Indels 5; Gaps 4;  
 1270 TGGTGTGTCATTTTAAAGTTTAAATAGAAACCTTTCTTTGTTAGTGCAGTCTCAACATC 1329  
 21 TGGTGTGTCATTTTAAAGTTTAAATAGAAACCTTCTTTGTTAGTGCAGTCTCAACATC 79  
 QY 1330 CCAAGGCACTGTAATGTATGACCGGCGGGTGTTCATGAGAGGCTCCAGTATGATCTTA 1389  
 DB 80 CCAAGGCACTGTAATGTATGACCGGCGGGTGTTCATGAGAGGCTCCAGTATGATCTTA 139  
 QY 1390 CATTCTAGTACGAGCTGTAATGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 1447  
 DB 140 CATTCTAGTACGAGCTGTAATGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 149  
 QY 1448 GCTCTGCGGATCGAGAGCTCTCTTCTTCTGAGCCCGGTGTCAGATGAGCTTATTTATGCT 1507  
 DB 200 GCTCTGCGGATCGAGAGCTCTCTTCTTCTGAGCCCGGTGTCAGATGAGCTTATTTATGCT 259  
 QY 1508 ATTATATGTAAATGCCACTGAAAGCTTAAAGTCTTAACTCTGGAATCCCAACACCAAGTT 1567  
 DB 260 ATTATATGTAAATGCCACTGAAAGCTTAAAGTCTTAACTCTGGAATCCCAACACCAAGTT 319  
 QY 1568 CTTCAAGGACTGCTGAGGAGCTGCTTATGAGGCTTCTTCTGAGGCTTCTTCTGAGGCTTCT 1627  
 DB 320 CTTCAAGGACTGCTGAGGAGCTGCTTATGAGGCTTCTTCTGAGGCTTCTTCTGAGGCTTCT 379  
 QY 1628 GGTTCACGACCCGACATGTGACATGAGGACATGATGCCGAAACCAACCCAGACACAC 1687  
 DB 380 GGTTCACGACCCGACATGTGACATGAGGACATGATGCCGAAACCAACCCAGACACAC 438

QY 1688 ATGCTCCATGTCAAGTGTGATGAGAGCAACTGCTCCCAAGCTGCTGCTCAGAGAG 1747  
 DB 439 ATGCTCCATGTCAAGTGTGATGAGAGCAACTGCTCCCAAGCTGCTGCTCAGAGAG 498  
 QY 1748 GTGTGAGTCTTCAAGTGTGCTGAGGAGGAGAGAGAGTGAAGTGTGCTGCTCTTTTAA 1807  
 DB 499 GTGTGAGTCTTCAAGTGTGCTGAGGAGGAGAGAGAGTGAAGTGTGCTGCTCTTTTAA 558  
 QY 1808 ATGTGCTTGAAGCTTTTAAAGTTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 1867  
 DB 559 ATGTGCTTGAAGCTTTTAAAGTTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 618  
 QY 1868 CACCTGAGGAGTGGCCCTTAAAGTGTGCTGAGTGAATCAAGCTTGTGACTGCGACAG 1927  
 DB 619 CACCTGAGGAGTGGCCCTTAAAGTGTGCTGAGTGAATCAAGCTTGTGACTGCGACAG 678  
 QY 1928 TCAGACTCCACCCAGTATGCTTTTGTCTTCTGCTTCTTCTTCTTCTTCTTCTTCTTCT 1987  
 DB 679 TCAGACTCCACCCAGTATGCTTTTGTCTTCTGCTTCTTCTTCTTCTTCTTCTTCTTCT 738  
 QY 1988 TGAGCTTAAGTGTATTTTATATGCTTCAATATCAAGTATGTAATGTAATGTAATGTA 2047  
 DB 739 TGAGCTTAAGTGTATTTTATATGCTTCAATATCAAGTATGTAATGTAATGTAATGTA 797  
 QY 2048 AAA 2050  
 DB 798 AAA 800

RESULT 15  
 AY403475  
 LOCUS  
 DEFINITION Mus musculus CDCA4 gene, VIRTUAL TRANSCRIPT, partial sequence.  
 AY403475  
 VERSION AY403475.1 GI:39759458  
 KEYWORDS GSS.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 714)  
 Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejarival,A.,  
 Todd,M.A., Tanendaaun,D.M., Civallo,D.R., Lu,F., Murphy,B.,  
 Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Snihsky,J.J.,  
 Adams,M.D. and Cargill,M.  
 Interfing nonneutral evolution from human-chimp-mouse orthologous  
 gene trios  
 Science 302 (5652), 1960-1963 (2003)  
 JOURNAL  
 PUBMED 14671302  
 REFERENCE 2 (bases 1 to 714)  
 AUTHORS Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejarival,A.,  
 Todd,M.A., Tanendaaun,D.M., Civallo,D.R., Lu,F., Murphy,B.,  
 Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Snihsky,J.J.,  
 Adams,M.D. and Cargill,M.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
 Rockville, MD 20850, USA  
 COMMENT This sequence was made by sequencing genomic exons and ordering  
 them based on alignment.  
 FEATURES  
 source  
 location/Qualifiers  
 1..714  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 <1..>714  
 /gene="CDCA4"  
 /locus\_tag="HCLM1569"

ORIGIN  
 Query Match 34.3%; Score 714; DB 29; Length 714;  
 Best Local Similarity 100.0%; Pred. No. 5.4e-132;  
 Matches 714; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 191 ATGTTGCTAGAGGCTGAAGAGAAATATGTGACCAAGAGAGAGTAGAGGTTTT 250
    |||||||
Db 1 ATGTTTCTAGAGGGCTGAAGAGAAATATGTGACCAAGAGAGAGTAGAGGTTTT 60
    |||||||
Qy 251 GGCAGTGTCCCTTCTTAAGCTTGACGACAGTCACTCTTGACATGTCCCTTGTCAAG 310
    |||||||
Db 61 GGCAGTGTCCCTTCTTAAGCTTGACGACAGTCACTCTTGACATGTCCCTTGTCAAG 120
    |||||||
Qy 311 CTCAGCTCTGTCAATGTGAGACCAATCTTGCCGCTGGTCTTCATGCGCAAC 370
    |||||||
Db 121 CTCAGCTCTGTCAATGTGAGACCAATCTTGCCGCTGGTCTTCATGCGCAAC 180
    |||||||
Qy 371 ACAGTCCGAGATCCGAGAGAAATGAGCAGAGTGTGTGSCATGGATGGCAACC 430
    |||||||
Db 181 ACAGTCCGAGATCCGAGAGAAATGAGCAGAGTGTGTGSCATGGATGGCAACC 240
    |||||||
Qy 431 CAGAAATGATAGCGGGCACCAGTTGAACGCTGTGTCAACAGATCTGTGTGTACA 490
    |||||||
Db 241 CAGAAATGATAGCGGGCACCAGTTGAACGCTGTGTCAACAGATCTGTGTGTACA 300
    |||||||
Qy 491 GTAGGGGAGCTGAGAGAGACCTGTCTGTGAATGGAATGCTCCCTTGCAAAAC 550
    |||||||
Db 301 GTAGGGGAGCTGAGAGAGACCTGTCTGTGAATGGAATGCTCCCTTGCAAAAC 360
    |||||||
Qy 551 TCGGTTTCCGAGCTCCCATCGTTGGCTCAGACCAAGGCAAGAAACCTCAGAGCAGC 610
    |||||||
Db 361 TCGGTTTCCGAGCTCCCATCGTTGGCTCAGACCAAGGCAAGAAACCTCAGAGCAGC 420
    |||||||
Qy 611 CTCGGAGATGAGCAGCCCAAGAAACAGGGGAAGTTTCAGAAATGACTGACACAG 670
    |||||||
Db 421 CTCGGAGATGAGCAGCCCAAGAAACAGGGGAAGTTTCAGAAATGACTGACACAG 480
    |||||||
Qy 671 ATATTGAGACCTTGAGAACAAAACTCAAGTTCAGTGAGGAATCTTCTCAGATGTG 730
    |||||||
Db 481 ATATTGAGACCTTGAGAACAAAACTCAAGTTCAGTGAGGAATCTTCTCAGATGTG 540
    |||||||
Qy 731 GACAGCTCCTACTATGACCTGAGCAGAGTCTAACAGGAATGATGAGTGGGACCAAGTCC 790
    |||||||
Db 541 GACAGCTCCTACTATGACCTGAGCAGAGTCTAACAGGAATGATGAGTGGGACCAAGTCC 600
    |||||||
Qy 791 AGTCTCTGCAATGAGGCTTGAAGGCTTGTGTCAGACCAACCCCTCCTCCAGTTCCACTTGC 850
    |||||||
Db 601 AGTCTCTGCAATGAGGCTTGAAGGCTTGTGTCAGACCAACCCCTCCTCCAGTTCCACTTGC 660
    |||||||
Qy 851 AAGTCTGACCTGCTGAGCTGACCAATGTGTAGAGATTCTGTGTGAGACCTGA 904
    |||||||
Db 661 AAGTCTGACCTGCTGAGCTGACCAATGTGTAGAGATTCTGTGTGAGACCTGA 714
    |||||||

```

Search completed: September 10, 2004, 14:17:08  
 Job time : 5672 secs

**THIS PAGE BLANK (USPTO)**